TECHNISCHE
UNIVERSITÄT
DARMSTADT

"From the bottom up - A systematic study of structure based coarse graining approaches"

Vom Fachbereich Chemie der Technischen Universität Darmstadt

zur Erlangung des Grades Doctor rerum naturalium
(Dr. rer. nat.)

Dissertation
von David Rosenberger

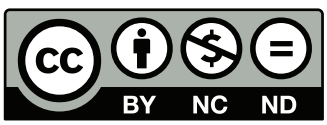
Erstgutachter/in: Prof. Dr. Nico van der Vegt
Zweitgutachter/in: Prof. Dr. Martin Hanke-Bourgeois

Darmstadt 2019

Tag der Einreichung: 28. Februar 2019

Tag der mündlichen Prüfung: 24.April 2019

Jahr der Veröffentlichung auf TUprints: 2019
URN: urn:nbn:de:tuda-tuprints-85094



Erklärung zur Selbstständigkeit und Promotionsversuch

Darmstadt, den

Ich erkläre hiermit, dass ich meine Dissertation selbstständig und nur mit den angegebenen Hilfsmitteln angefertigt und noch keinen Promotionsversuch unternommen habe.

(Rosenberger, David)

Erklärung der Übereinstimmung

Darmstadt, den

Ich erkläre hiermit, dass die elektronische Version der Doktorarbeit mit der schriftlichen Version übereinstimmt. Die elektronische Version liegt dem Prüfungssekretariat vor.

(Rosenberger, David)

Erklärung zum Eigenanteil an den Veröffentlichungen

Darmstadt, den

Die vorliegende Dissertation wurde unter der Leitung von Herrn Prof. Dr. Nico van der Vegt in der Zeit vom 02. Februar 2015 bis zum 28. Februar 2019 im Fachbereich Chemie der Technischen Universität Darmstadt durchgeführt.

Teile dieser Arbeit sind bereits veröffentlicht oder werden in Kürze veröffentlicht.

Mein Anteil an der folgenden Veröffentlichung beträgt 100 %.
Rosenberger, D.; Hanke, M.; van der Vegt, N. F. A., Comparison of iterative inverse coarse-graining methods,*Eur. Phys. J. Spec. Top. 225, 1323-1345 (2016)*

Mein Anteil an der folgenden Veröffentlichung beträgt 100 %.
Rosenberger, D.; van der Vegt, N. F. A., Addressing the temperature transferability of structure based coarse graining models, *Phys. Chem. Chem. Phys. 20, 6617-6628 (2018)*

Mein Anteil an der folgenden Veröffentlichung beträgt 80 %.
Rosenberger, D.; Sanyal, T.; Shell, M. Scott; van der Vegt, N. F. A., Transferability of local density assisted implicit solvation models for homogeneous fluid mixtures, *Article ASAP in J. Chem. Theory Comput., DOI:10.1021/acs.jctc.8b01170 (2019)*

Mein Anteil an der folgenden Veröffentlichung beträgt 100 %.
Rosenberger, D.; van der Vegt, N. F. A., The relative entropy indicates an ideal concentration for structure based coarse graining of binary mixtures, *just accepted in Phys. Rev. E*, am 06.05.2019.

(Rosenberger, David)

Erklärung zur Begutachtung der Veröffentlichungen

Referent: Prof. Dr. Nico van der Vegt
Korreferent: Prof. Dr. Martin Hanke-Bourgeois

Weder Referent (Prof. Dr. Nico van der Vegt) noch Korreferent (Prof. Dr. Martin Hanke-Bourgeois) waren an der Begutachtung nachstehender Veröffentlichungen beteiligt:

(1): **Rosenberger, D.**; Hanke, M.; van der Vegt, N. F. A., Comparison of iterative inverse coarse-graining methods, *Eur. Phys. J. Spec. Top. 225, 1323-1345 (2016)*

(2): **Rosenberger, D.**; van der Vegt, N. F. A., Addressing the temperature transferability of structure based coarse graining models, *Phys. Chem. Chem. Phys. 20, 6617-6628 (2018)*

(3): **Rosenberger, D.**; Sanyal, T.; Shell, M. S.; van der Vegt, N. F. A., Transferability of local density assisted implicit solvation models for homogeneous fluid mixtures, *Article ASAP in J. Chem. Theory Comput., DOI:10.1021/acs.jctc.8b01170 (2019)*

(4): **Rosenberger, D.**; van der Vegt, N. F. A., The relative entropy indicates an ideal concentration for structure based coarse graining of binary mixtures, *just accepted in Phys. Rev. E*

Datum:

Referent:
(Prof. Dr. Nico van der Vegt)

Korreferent
(Prof. Dr. Martin Hanke-Bourgeois)

Acknowledgments

A journey comes to an end. A journey which would have been impossible to start and to end without the contributions of many people. I would like to take the following lines to say thank you. Foremost to my supervisor Nico van der Vegt. His guidance, encouragement and trust was key to the success of this work. Getting the chance to pursue my own ideas during my Ph.D. and at the same time getting full support and finding always an open door for discussion made my doctoral studies a fruitful and enjoyable time. I would also like to thank Prof. Martin Hanke-Bourgeois from the mathematics department of the Johannes Gutenberg University in Mainz and his co-workers Fabrice Delbary and Dimitri Ivanizki. Their patience in teaching me basic numerics made a huge impact on this work. I'm thankful to Prof. Florian Müller-Plathe from the Theoretical Physical Chemistry Group at TU Darmstadt and Prof. Friederike Schmid from the Statistical Physics and Soft Matter Theory group at the Johannes Gutenberg University in Mainz for being part of my evaluation committee. Further, I'm grateful to Prof. M. Scott Shell and his student Tanmoy Sanyal from the University of California Santa Barbara from whom I learned a lot during my 3 months of exchange stay in the US. I also appreciate all the insights I got from Prof. Will Noid and his student Nicholas Dunn from Penn State University. Further, I would like to thank Christoph Scherer, Joe Rudzinski and Denis Andrienko from the Max Planck Institute for Polymer Research in Mainz for helping me making the first steps in a new research field and show continuous interest in the progress I made. I'm also thankful to Christoph Junghans from the Los Alamos National Laboratory for supporting me, especially in software related questions.
Of course I'm also thankful to all members of the Computational Physical Chemistry group at TU Darmstadt. Not only for the past 4 years of my Ph.D., but also for the time since I joined the group as a Bachelor's student back in 2011. Working in this group has always been a pleasure. I enjoyed the balance between hard-working phases and the socially relaxing atmosphere created by all group members a lot. Part of the group have always been helpful people, who made my life easier. So, thank you Imke, Marianne and Ellen.
A Ph.D. can never be successful without finding the right balance between work and leisure time. I want to thank the people whom I met over the past 10 years in Darmstadt and who helped me finding this balance. So, thank you Gregor, Henner, Kai, Velten, Valentina,Vikram, Francisco, Emiliano, Pritam, Pim, Armin, Aki, Jonas, Martin, Steffen, Thomas, Ömer, Divya, Tobias, Philipp, Vincent, Apoorv and Sawmi.
Last, but not least I want to thank my family for their support through the past years. Although not really understanding what I actually do, they always lent me an ear when I needed it. It has always been a good feeling to know that I can rely on you. Speaking of which, thank you Ekaterine. Thank you for being there for me for the last three years. Thank you for being understanding when I worked long hours or when I was hesitant in taking vacation. Thank you for cheering me up after a long a day at work. Thank you for being a part of my life.

Summary

Computer simulations of soft matter require a compromise to be made between the computational efficiency and the resolution of a model studied. Highly resolved models can give insights into the interactions between individual atoms in a soft material. But, since these atomistic models are computational expensive, they are limited to small length scales and short time scales, which makes it difficult to compare simulation results with the ones from experiments in the laboratory. On the other hand, continuum models enable the study of soft matter at larger length and longer time scales and are computationally less expensive, but they rather focus on macroscopic properties than on their atomistic origin. A possible way to bridge the gap between these scales is to perform simulations at an intermediate level of resolution. The problem, which exists at this mesoscopic scale, is the lack of accurate models. Hence, new ones have to be built. The process to construct mesoscopic models based on information from the atomistic scale is commonly referred to as bottom-up coarse graining. Bottom-up coarse graining describes the process of lowering the resolution of a atomistic model to make it applicable at larger length and time scales. The major goal of this Ph.D. thesis is to increase the knowledge on so called structure-based bottom-up coarse graining techniques.These methods enable the derivation of coarse grained (CG) models, which accurately reproduce the structure of an atomistic or fine grained (FG) model at the mesoscopic scale. The shortcomings of different structure-based methods are carefully analyzed and new approaches to overcome them are presented.

The thesis is structured in the following way. We start with a brief introduction into the field of computer simulations of soft matter systems and coarse graining. In Chapter 2 the necessary theoretical background is introduced. This chapter contains a basic introduction into molecular dynamics (MD) simulations in different thermodynamic ensembles and a more elaborate discussion on bottom-up coarse graining, and in particular on structure based coarse graining.

Chapter 3 contains the main results of this work based on 4 peer-reviewed articles. In chapter 3.1 a systematic comparison between different structure based coarse graining methods is presented. This study facilitates the choice of a coarse graining method based on systematic criteria like convergence, numerical stability and theoretical assumptions, which underlie each method. In addition, we examine the performance of multiple CG models to reproduce properties different than the structure, which is known as representability. We observe that matching structure and thermodynamics simultaneously is difficult. This indicates that reproducing the structure is not sufficient to represent a FG model at a CG level. This aspect is further discussed in chapter 3.2.

One way to overcome this shortcoming is to introduce additional energetic contributions to the CG model. In chapters 3.2, 3.3 and 3.4 we investigate the role of local density dependent and volume dependent potentials on the representability as well as on the concentration and temperature transferability of the derived CG models. Transferability means the model is applicable at state points different than the ones chosen for parametrization. In addition, in chapter 3.2 we propose a novel idea to improve the concentration transferability of structure based coarse graining models. The novelty lies in the identification of an ideal state point to parametrize transferable CG model based on the relative entropy.

To shortly summarize the main findings: Matching structures might be a poor choice to develop CG models, which should additionally represent thermodynamic properties and which are transferable to different concentrations or temperatures. Nevertheless, this shortcoming can be compensated for by either systematically selecting the state point for model parametrization or by augmenting the model with additional energetic contributions. A more detailed summary of the insights achieved in this thesis is given in the final chapter along with an outlook into the future of bottom-up coarse graining.

Zusammenfassung

Computersimulationen weicher Materie unterliegen einem Kompromiss zwischen dem Rechenaufwand und der Auflösung des verwendeten Modells. Hoch aufgelöste Modelle erlauben es die Wechselwirkungen zwischen einzelnen Atomen in weichen Materialen zu untersuchen. Jedoch sind diese Modelle in ihre Anwendbarkeit limitiert, weil sie aufgrund ihres hohen Rechenaufwandes nur auf kurzen Längen- und Zeitskalen verwendet werden können. Dies erschwert den Vergleich von Simulationsergebnissen mit den Ergebnissen, welche aus Laborexperimenten erhalten werden. Verringert man die Auflösung der Modelle, verkleinert man zwar den Rechenaufwand und überwindet die Zeit- und Längenskalenbeschränkung, jedoch verliert man gleichzeitig Information über die atomaren Wechselwirkungen. Um die Lücke zwischen diesen Skalen zu schliessen, besteht die Möglichkeit atomistische Modelle zu vergröbern. Findet die Vergröberung basierend auf Information höher aufgelöster Modelle statt, bezeichnet man die Vergröberungsmethoden auch als von unten nach oben Vergröberungstechniken. Im Rahmen dieser Doktorarbeit beschäftige ich mich mit Vergröberungstechniken, welche die Struktur eines hoch aufgelösten Modelles auf der vergröberten Ebene exakt wiedergeben. Die Schwächen dieser strukturbasierten Methoden werden analysiert und neue Lösungswege diese zu verbessern werden vorgestellt.
Die Arbeit ist wie folgt gegliedert. In der Einleitung wird ein Überblick über Computersimulationen von weicher Materie präsentiert zusammen mit einer allgemeinen Vorstellung von verschiedenen Vergröberungstechniken. In Kapitel 2 werden die wichtigsten theoretischen Grundlagen erläutert und die hier behandelten Probleme ausführlich dargestellt.
Kapitel 3 beinhaltet die Ergebnisse dieser Doktorarbeit, welche in Form von vier wissenschaftlich begutachteten Veröffentlichungen vorliegen. In Kapitel 3.1 werden verschiedene strukturbasierte Vergröberungstechniken systematisch miteinander verglichen. Dies ermöglicht es objektive Entscheidungskriterien zu ermitteln, aufgrund derer man sich für eine Methode entscheiden kann. Darüber hinaus wird die Fähigkeit der Modelle untersucht thermodynamische Größen zusätzlich zur Struktur zu reproduzieren. Dies gestaltet sich als schwierig, was darauf hindeutet, dass die Reproduzierbarkeit der Struktur nicht ausreichend ist um ein Referenzmodell in vergröberter Darstellung zu beschreiben. Dieser Aspekt wird in Kapitel 3.2 weiter beleuchtet.
Um diese Schwachstelle zu umgehen kann man zusätzlich Energiebeiträge in das vergröberte Modell einführen. In den Kapiteln 3.2, 3.3 und 3.4, werden die Auswirkungen dieser Energiebeiträge, welche zum einen lokale Dichteeffekte und zum anderen Volumenfluktuationen berücksichtigen, untersucht. Mit diesen zusätzlichen Beiträgen ist es sowohl möglich thermdoynamische Größen zu reproduzieren, als auch die Transferierbarkeit der entwickelten Modelle zu erhöhen. Transferierbarkeit beschreibt die Fähigkeit der Modelle Systeme an Zustandspunkten zu beschreiben, welche nicht Teil der Parametrisierung waren.
Zusätzlich präsentieren wir in Kapitel 3.2 einen neuen Ansatz um die Transferierbarkeit vergröberter Modelle zu erhöhen. Dieser neue Ansatz beruht auf der Identifikation eines idealen Zustandpunktes zur Modellparametrisierung, welcher mit Hilfe der relativen Entropie detektiert werden kann.
Die Ergebnisse aus Kapitel 3 werden schließlich in Kapitel 4 zusammengefasst. Kurz an dieser Stelle: die Reproduzierbarkeit von Struktur auf vergröberter Ebene stellt eine zu grobe Annäherung dar um das komplette Referenzsystem auf einer vergröberten Ebene zu simulieren. Jedoch kann man dieses verbessern, sei es durch kluges Auswählen eines Zustandspunktes oder durch zusätzliche Energiebeiträge. Zusätzlich beinhaltet dieses letzte Kapitel noch einen Ausblick auf zukünftige Herausforderungen im Bereich der von unten nach oben Vergröberungstechniken.

Contents

Introduction

Soft condensed matter, like polymers or complex liquids, is characterized by interaction energies of the order of k_BT at room temperature T, where k_B is the Boltzmann constant.[1–4] Thus, thermal fluctuations can induce structural and conformational changes in soft materials, which makes these systems highly flexible. As a consequence, it can take several seconds for soft materials to reach an equilibrium state at macroscopic length scales (millimeter to meter). This makes it difficult to study soft matter with the aid of computer simulations. But, computer simulations enable to study soft matter at resolutions difficult to access with common experimental techniques.[5–9] The problem with computer simulations arises from the fact that the common methods, classical molecular dynamics (MD)[10] and Monte-Carlo (MC)[11] simulations, are limited to shorter length and time scales[12] than those required to account for the equilibration of soft matter at macroscopic length scales. Therefore, there is a necessity to reach larger length and time scales with computer simulations on the one hand. On the other hand, one has to simultaneously account for length and time scales at which microscopic changes occur (picometer, femtoseconds), for example the formation of hydrogen bonds. Hence, modeling of soft materials is a multiscale problem,[13] which is illustrated in figure 1.

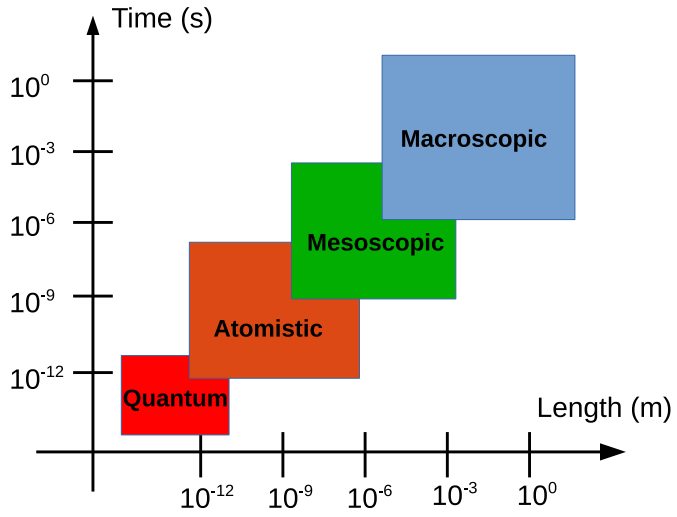


Figure 1: Illustration of the multiscale problem in computer simulations of soft matter systems.

Each of the scales involved can be addressed with a specific simulation approach. Quantum mechanical approaches address the smallest length and shortest time scale behavior in soft materials. MD, MC or dissipative particle dynamics (DPD)[14,15] simulations focus on the other hand on the atomistic and mesoscopic scale or resolution. The macroscopic level can be described with continuum models like finite element methods. All these approaches are based on the assumption that each scale can be addressed indpendent of the others. Multiresolution approaches improve this by bridging between two or more resolutions in one modeling approach: The class of quantum mechanics/statistical mechanics (QM/MM) methods bridges between the quantum and atomistic length and time scales.[16–20] Adaptive resolution techniques also enable the coupling between these two scales.[21] Further, they can be applied to combine the atomistic and mesoscopic length and time scales.[22–25] Smoothed dissipative particle dynamics,[26–30] smoothed particle hydrodynamics[31,32] or methods like self-consistent field modeling[33–35]

or Lattice Boltzmann approaches[36–39] can fill the gap between the atomistic/mescoscopic scale and the continuum level. Although the methods have become more applicable over the last years, their computational efficiency is - in many cases - still limited by the lowest resolution present in the system. A different approach to solve the multiscale problem is to generate hierarchical models, which are constructed based on information obtained from a different resolution. The process of going from a highly detailed model to a lower one is commonly referred to as coarse graining and is the main topic addressed in this work.

Coarse graining is basically a two step process. In the first step, a high resolution, or fine grained (FG), model is projected onto a configuration space lower in resolution, the so called coarse grained (CG) configuration space. For example, the projection from the atomistic to the mesoscopic scale means that individual atoms are grouped together as effective particles larger in size. This consequently leads to a decrease in the total number of particles present in the system. Inherent to the reduction of the total particle number, is a loss in the number of degrees of freedom (DOFs) available to the system.[40] Thus, the way the particles are grouped together, also known as the mapping scheme, determines which DOFs are lost. A optimal mapping scheme should be chosen in a way that the DOFs irrelevant to the physical problem addressed are lost and the important ones are kept. The problem is there is no systematic way how to determine an optimal mapping scheme in a practical way and its selection heavily relies on the modeler's chemical intuition.[41–43]
The second step is the generation of the effective potentials to describe the interactions between the CG particles. The optimal solution to this problem is given by the so called many-body potential of mean force (m-PMF) $W(\mathbf{R})$.[40]

$$W(\mathbf{R}) = -k_B T \ln \int_V d\mathbf{r} \exp[-\beta U_{FG}(\mathbf{r})] \delta(\mathbf{R} - \mathbf{M}(\mathbf{r})) \,, \tag{1}$$

where $\mathbf{R}$ is the CG configuration obtained from the mapping of the FG configuration, $\mathbf{r}$. The mapping is defined through the mapping operator $\mathbf{M}(\mathbf{r})$, k_B is the Boltzmann constant, T and V are the temperature and the volume of the system respectively, β is $1/k_B T$, $U_{FG}(\mathbf{r})$ is the total potential energy of the FG system, and the Dirac δ distribution filters those FG configuration which map to the same CG configuration. $W(\mathbf{R})$ guarantees that all mapped configurations occur with the same probability as they would in the FG system. Unfortunately, to evaluate the integral in Eq. 1 a highly multibody problem has to be solved, which is practically infeasible. Therefore, only approximate solutions of Eq. 1 can practically be obtained.[44] These solutions approximate the m-PMF via a CG potential, $U_{CG}(\mathbf{R})$, of the generalized form:

$$W(\mathbf{R}) \approx U_{CG}(\mathbf{R}) = \sum_{\zeta} \sum_{\lambda} U_{\zeta}(\psi_{\zeta}(\mathbf{R}_{\lambda})) \,, \tag{2}$$

where ζ describes a specific type of bonded or non-bonded interaction which is modeled via the corresponding potential U_{ζ}, which depends on a scalar variable ψ_{ζ}, which is a function of the coordinates $\mathbf{R}_{\lambda}$ for a set λ of CG particles.[40] To give an example: For a non-bonded interaction between a pair of particles, ζ represents the non-bonded pair interaction, U_{ζ} is the corresponding potential, ψ_{ζ} is the pair distance between the particles and $\mathbf{R}_{\lambda}$ are the coordinates for a pair λ of CG particles. In order to obtain $U_{CG}(\mathbf{R})$ one can follow two different approaches.
The first one is the top down or inverse coarse graining approach.[45,46] Here, CG models are parametrized based on a fixed functional form to describe the effective particle-particle interactions. The functional

form depends on parameters, which are determined to accurately reproduce experimental, thermodynamic and/or structural quantities.[47–49] Following this rather generic approach, one can study physical processes in various systems, which are accurately described by the chosen functional form.[50,51] Inverse CG models have been succesfully applied to study the self assembly of amphiphilic molecules,[52–54] in particular membranes,[55–57] or the phase behavior of polymers.[58–60]
Although the popular Martini model,[61–64] among others,[65,66], can introduce some chemical specificity to top down models, their rather generic character is a clear disadvantage, if one is interested in a specific system or system specific effects, like conformational flexibility and stability of native protein structures under different environmental conditions.[67–69]

In order to generate more specific CG models and to learn more about the microscopic driving forces, bottom up or systematic coarse graining methods can be applied. Here, CG models are derived systematically from a specific FG model of a system. Systematic in the context of coarse graining means that a CG models is derived based on specific information extracted from the FG model. Multiple systematic or bottom-up coarse graining methods exist to generate models, which enable the study of processes like the mixing behavior of benzene and water.[70,71] or the assembly behavior of proteins[72–75] and polymer melts[76–78] at a coarse grained level. Bottom-up coarse graining methods can be generally classified into two different groups, namely the derived coarse graining methods and the inverse methods.
Derived coarse graining methods generate CG models based on the computation of effective pair potentials at a FG level from the direct interactions between pairs of groups of atoms embedded in the effective particles. Methods like the pair potential of mean force approach,[79] effective force coarse graining[80] or the conditional reversible work method[81] belong to this class. Application of these methods leads to CG models, which can qualitatively describe the thermal behavior of alkanes[82] or ionic liquids[80] without representing the FG structure in a quantitative manner.
A better quantitative description of a FG system is obtained by application of inverse coarse graining methods. These methods try to find an approximation to the m-PMF (see Eqs. 1 and 2), by minimizing the difference between the FG and the CG system with respect to one or multiple target properties. The multiscale coarse graining (MS-CG) approach uses the particle forces as a target.[83] Here, the difference in the particle pair forces between the FG and CG system are minimized following a variational principle,[84,85] which is why it is also known as force matching.[86] MS-CG has been successfully applied to derive CG models for peptides[87] or lipid bilayers.[88] The generalized-Yvon-Born-Green (g-YBG) method[89,90] is also based on a variational principle and generates MS-CG potentials without explicitly accounting for forces. Instead, structural differences between the CG and FG system are minimized following the Yvon-Born-Green-hierarchy, i.e information on two and three body structural correlations are taken into account.[91] This approach has led to accurate CG models, where simple pairwise additive forces are derived from higher order structural correlations. G-YBG models can quantitatively describe the Honeycutt-Thirumulai model peptide,[89] liquid alkanes[92] or polyethylene based ionomers.[93]
CG models derived on the basis of variational principles are capable to quantitatively describe the average force along a single degree of freedom, but they do not quantitatively reproduce cross correlations between different DOFs.[91] As a consequence, these potentials fail to accurately represent the distribution functions of each degree of freedom (DOF) as predicted by the FG system.[40] In order to improve this representability problem an iterative correction can be applied. The iterative version of the g-YBG approach (iter-g-YBG) optimizes the representability of the FG pair distribution function by solving the variational principle in a self-consistent manner.[94] Similarily, the iterative Boltzmann inversion (IBI)[95,96] or the inverse Monte Carlo (IMC) approach[97] iteratively update the pair PMF between CG particles, until the difference between the radial distribution function (RDF) of the FG and CG system is minimized.

Thus, IBI and IMC CG models can quantitatively describe the pair structure of the FG model at a CG level. Another iterative approach is the relative entropy optimization introduced by Shell.[98] Here, the information loss defined in terms of the relative entropy is minimized. This quantity can be derived on the basis of the Kullback-Leibler divergence.[99,100] Depending on which minimization scheme is chosen, relative entropy optimization can either result in IBI/IMC or MS-CG potentials, without explicitly including information about the pair structure or the forces.[101]
Despite the success of inverse coarse graining methods to generate effective potentials, which quantitatively describe the target property, they lack the ability to represent other quantities with the same accuracy.[102,103] For example, IBI and IMC models cannot accurately model the pair structure and the pressure simultaneously, since the latter one is mostly determined by the long range attraction, whereas the structure is mainly determined by the short range repulsion between two particles.[104] Several approaches exist to compensate for this effect, among which the most famous is the addition of a ramp potential. This additional contribution alters the effective interaction potential to reproduce both, pressure and RDF, after the structure of the FG system is quantitatively matched.[96]
Further, most approximate solutions of the m-PMF derived with inverse bottom-up coarse graining methods are state point dependent. Thus, they are barely transferable to state points different from the one chosen during parameterization.[105–107] Although parametrization approaches, which include the information of multiple state points[108,109] or which apply ramp potentials to reproduce thermodynamic properties like Kirkwood-Buff integrals,[110] have significantly improved bottom-up CG models, their transferability is limited to a certain range. Further, there is no obvious connection between accurate representability and transferability, what makes it difficult to assess, which DOFs have to be included in the CG model or how to compensate for their loss to achieve accurate representability and transferability.

In this thesis, both of the issues - representability and transferability - are addressed in the context of structure based coarse graining, in particular the iterative inverse methods IBI and IMC. The methods are chosen because an accurate description of the structure is necessary in order to study processes at a CG level, which are characterized by structural changes. The main goal of this work is to gain deeper knowledge on the methods itself and how to overcome their shortcomings in a systematic way. The generated knowledge provides guidelines to develop CG models for an actual application at a later stage. Therefore, a systematic comparison of different structure based coarse graining methods is presented in chapter 3.1. Here, we investigate the convergence of a variety of IBI-type methods to a known solution to the inverse problem in comparison with IMC. This facilitates the choice of a coarse graining method based on systematic criteria like convergence, numerical stability and theoretical assumptions, which underlie each method. In addition, we examine the representability of the derived models with respect to properties different than the liquid structure at pair level. We observe that matching structure and thermodynamics remains difficult, despite the different advantages and disadvantages of the methods elaborated.
Hence, we have to find ways to overcome the still existing shortcomings. One way to achieve this is to introduce additional energetic contributions to the CG model, which account for the loss of DOFs. In chapters 3.2, 3.3 and 3.4 we investigate the role of local density dependent and volume dependent potentials on the representability as well as on the concentration and temperature transferability of the derived CG models. What is more, in chapter 3.3 we also prove the analytical and numerical equivalence between relative entropy optimization and IMC. This demonstrates the possibility to derive structurally accurate CG models without explicitly selecting the liquid pair structure as a target for the parametrization.
In addition, in chapter 3.2 we propose a novel idea to improve the concentration transferability of

structure based coarse graining models. The novelty lies in the identification of an ideal state point to parametrize transferable CG model based on the relative entropy.
Further, in chapter 3.4 we introduce a computationally cheap approach to parametrize a volume dependent potential, which follows a simple linear regression scheme. This allows to overcome several shortcomings of structure based CG models, which have not been resolved before.
The thesis is structured as follows. Chapter 2 introduces the theoretical background relevant for this work. A basic introduction into MD simulations and bottom-up coarse graining is followed by a more detailed discussion on the problems addressed in this thesis. Chapter 3 contains the main results based on 4 peer-reviewed articles. In chapter 4 we finally summarize the main findings of chapter 3 and put them in context of the main problems discussed in this work along with an outlook on remaining challenges and on the future of the field of bottom-up coarse graining.

Theoretical background

2.1 Molecular Dynamics

As stated in the introduction, MD simulations enable the study of processes on atomistic length (nm) and time (ns) scales. This is possible by integrating Newton's equations of motions (see Eq. 3) for each atom present in a N- particle system, where each particle has an assigned mass, charge and diameter.[111]

$$\mathbf{F_i} = m_i \frac{d^2\mathbf{r}_i}{dt^2} \ , \tag{3}$$

where m_i is the mass of particle i, $\mathbf{r}_i$ is the position vector, t is the time and $\mathbf{F}_i$ is the force acting on the $i-$th particle in the system. Since an analytical solution of Eq. 3 is almost impossible for large N, a numerical integration scheme is commonly applied to solve Newton's equations of motions. Subject to a set of initial positions, $\mathbf{r}^N = \{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N\}$, velocities, $\mathbf{v}^N = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ and forces, $\mathbf{F}_i$, Eq. 3 gets integrated n-times over all particles i. A time interval Δt, also known as time step, is between each integration step. The initial configuration for a system of interest can, for example, be generated from a random distribution of particles on a cubic lattice, where each lattice site can be occupied by a single particle only.[112] Initial velocities can be assigned to each particle according to the Maxwell-Boltzmann distribution,[113]

$$P(v) = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} v^2 \exp\left(-\frac{mv^2}{2k_B T} \right) \ , \tag{4}$$

where $P(v)$ is the probability distribution of the velocities v, m is the particle mass, k_B is the Boltzmann constant and T is the temperature of the system. The distribution of velocities should lead to an average kinetic energy, E_{kin}, which matches the desired temperature of the system according to

$$E_{kin} = \frac{3}{2} N k_B T \ , \tag{5}$$

where N is the number of particles in the system, k_B and T have the same meaning as before.[114] The forces, which act on each particle i are determined by the negative derivative of the potential energy $U(\mathbf{r}^N)$ with respect to the position of the particle, $\mathbf{r}_i$

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{r}^N)}{\partial \mathbf{r}_i} \ , \tag{6}$$

where $U(\mathbf{r}^N)$ is the potential energy of the system, which is defined by a set of parameters and a functional form commonly referred to as force field (FF).[113]
After initialization, the numerical integration is performed n- times using special integrator algorithms, which generate a trajectory, i.e. the particle's position as a function of the time t and the initial positions and velocities. Assuming ergodicity, which means that averaging over time is the same as averaging over the configuration ($\mathbf{r}^N$) and the momenta space ($\mathbf{p}^N$),[114]

$$\langle a \rangle = \frac{\int\int d\mathbf{r}^N d\mathbf{p}^N a \delta(H(\mathbf{r}^N, \mathbf{p}^N) - E)}{\int\int d\mathbf{r}^N d\mathbf{p}^N \delta(H(\mathbf{r}^N, \mathbf{p}^N) - E)} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau dt \ a(\mathbf{r}_t^N, \mathbf{p}_t^N) \ , \tag{7}$$

we can compute the average of a quantity a from a trajectory by the means of statistical thermodynamics, as long as a is a function of the configuration and momenta space. In Eq. 7 $\langle ... \rangle$ denotes the average of a quantity, E is the total energy of the system, H is the classical Hamiltonian (see Eq. 9 for definition) and all other variables have the same meaning as previously defined. The integral on the right hand side (RHS) in Eq. 7 assumes continuity in time t, but MD trajectories are discretized in time based on the chosen time step Δt. Therefore, the time average of a property $a(\mathbf{r}^N, \mathbf{p}^N)$ obtained from MD simulations is computed as[114]

$$\langle a \rangle = \frac{1}{M} \sum_{n=1}^{M} a(\mathbf{r}^N_{n\cdot\Delta t}, \mathbf{p}^N_{n\cdot\Delta t}) \, , \tag{8}$$

where M is the number of integration steps performed. To ensure that Eq. 7 and Eq. 8 are equal, a sufficient number of integration steps has to be performed.
In practice, several algorithm exist to numerically solve Newton's equations of motions, among which the two most common are the velocity verlet algorithm[115] and the leap-frog algorithm[116], where the latter one is explained more detailed in chapter 2.1.1.
Numerical integration of Eq. 3 generates an energy conserving ensemble, the so called microcanonical ensemble. A microcanonical ensemble is a thermodynamically isolated system, where there is no exchange of volume (V), number of particles (N) and energy (E) with the environment. The total energy of such a system is defined by means of the classical Hamiltonian[111]

$$H(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N) \, , \tag{9}$$

where m_i, $\mathbf{p}_i$ and $\mathbf{r}_i$ are the mass,the momentum and the position of a particle i and $U(\mathbf{r}^N)$ is the potential energy depending on the position of the N-particles $\mathbf{r}^N$. In order to relate the results of MD simulations to most macroscopic physical properties, the simulations have to be performed in either thermodynamically closed systems, like the canonical (constant N,V,T) or the isothermal-isobaric (constant N,P,T) ensemble, or thermodynamically open systems (constant μ,V,T). In different ensembles, the conserved quantities change and thus the equations of motions. This will be further discussed in the following sections.

2.1.1 Molecular dynamics in the canonical ensemble

In order to explain how MD simulations can be performed in the canonical ensemble, we recall Eq. 5. This equation defines the temperature of the system based on the kinetic energy.
Thus, to keep the temperature of the system constant at a desired value, algorithms like velocity rescaling[117] or weak coupling[118] rescale the particle velocities such that the kinetic energy of the system matches the desired temperature. But, despite guaranteeing the correct temperature, velocity rescaling[117] does not resemble the correct thermal fluctuations present in the canonical ensemble. Instead, this leads to the so-called isokinetic ensemble.[114] Further, the weak coupling method of Berendsen[118] indeed enables fluctuations around the desired temperature value, but these fluctuation do not correspond to those generated by a true canonical ensemble. A possible way to overcome this shortcoming is to apply the so-called Nosé-Hoover thermostat.[119,120] which introduces additional DOFs to the system's Hamiltonian (see Eq. 9). This changes the conserved energy to[114]

$$H_{NH}(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N) + \sum_{j=1}^{M} \frac{p_{\eta_j}^2}{2Q_j} + gk_BT\eta_1 + k_BT \sum_{j=2}^{M} \eta_j \, , \tag{10}$$

where η_j can be considered as an external heat bath, which acts on the system through M contributions and has a "mass" Q. In Eq. 10, g equals the number of degrees of freedom and all other variables have the same meaning as in Eq. 9. The additional terms in Eq. 10 have to be further included in Newton's equations of motions, whose exact explanation is beyond the scope of this work and can be found elsewhere.[112]

In the context of coarse graining it has to be taken into account that due to the loss of particles less collision events take place and less friction induced by the environment is present. Therefore, the application of the Nosé-Hoover thermostat is not as desirable. Langevin dynamics or stochastic dynamics introduces two additional terms to Newton's equations of motion to account for those effects.[121]

$$m_i \frac{d^2\mathbf{r}_i}{dt^2} = -m_i\gamma_i \frac{d\mathbf{r}_i}{dt} + \mathbf{F}_i(\mathbf{r}) + \mathring{\mathbf{R}}_i(t), \tag{11}$$

where γ is a friction coefficient and $\mathring{\mathbf{R}}_i(t)$ is a random force, which follows a Gaussian distribution with zero mean. The random force is required to satisfy the fluctuation dissipation theorem, which ensures that the temperature stays constant.[111]

$$\langle \mathring{\mathbf{R}}_i(t_0)\mathring{\mathbf{R}}_i(t)\rangle = 2m_i k_B T\,\delta(t-t_0) \tag{12}$$

The advantage of Langevin dynamics is it yields to the canonical ensemble without introducing additional equations of motions as compared to the Nosé-Hoover approach. The two additional terms in Eq. 11 can simply be implemented in the regular leap frog algorithm, which runs as follows:[122]

$$\mathbf{v}_i = \mathbf{v}_i\left(t - \frac{1}{2}\Delta t\right)\frac{\mathbf{F}_i(t)}{m_i}\Delta t\ , \tag{13}$$

$$\Delta\mathbf{v}_i = \frac{1-\alpha}{m_i\gamma}\mathbf{F}_i(t) + \sqrt{\frac{k_B T}{m_i}(1-\alpha^2)}\mathring{\mathbf{R}}_i\ , \tag{14}$$

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \left(\mathbf{v}_i + \frac{1}{2}\Delta\mathbf{v}_i\right)\Delta t\ , \tag{15}$$

$$\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \alpha\mathbf{v}_i(t-\Delta t) + \Delta\mathbf{v}_i\ . \tag{16}$$

First (Eq. 13), the particle velocities $\mathbf{v}_i$ are updated at every time step Δt based on the forces, which act on the particles at time t ($\mathbf{F}_i(t)$). The velocity update is computed based on a first order Taylor approximation, where all higher order terms are omitted. In Eq. 13 and in all following equations m_i is the particle mass. In a second step (Eq. 14), these velocities are reduced according to the additionally introduced friction and random noise term as defined in Eq. 11. In the third step (Eq. 15), the updated velocities are used to calculate the new positions of the particles at the full time step ($\mathbf{r}_i(t+\Delta t)$). Again, this update is derived based on a Taylor approximation where the higher order terms are ignored. In the fourth step (Eq. 16) the modified velocities are assigned to actual particle velocities. In Eqns. 14 and 16 α is defined as:

$$\alpha = \left(1 - \frac{\gamma\Delta t}{m_i}\right) \tag{17}$$

where all variables have the same meaning as previously defined. In the last step of the leap frog algorithm, the forces on each particle get updated based on the potential energy of the new configuration of the system according to Eq. 6. These steps are repeated n-times until the system has progressed in time as defined in the beginning of the integration process.

2.1.2 Molecular dynamics in the isothermal-isobaric ensemble

To perform MD simulations at a constant pressure P, the system has to be able to adjust its volume, such that the pressure inside the system equals the external pressure enforced by a barostat. In particular, this is important in order to investigate a system's reaction upon a change in temperature as discussed in chapters 3.2 and 3.4. As for the canonical ensemble, a correct isothermal-isobaric (NPT) ensemble cannot be achieved by scaling algorithms like the Berendsen barostat[118], although they guarantee constant pressure by rescaling the system's volume. For an exact simulation of the NPT ensemble, additional DOFs have to be added to the system's Hamiltonian. The conserved energy is then given by[114]

$$h_1(V^{-1/3}\mathbf{r}^N, V^{1/3}\mathbf{p}_i, V, \Pi) = \sum_{i=1}^{N} \frac{V^{-1/3}\mathbf{p}_i^2}{2m_i V^{2/3}} + U(V^{1/3}, V^{-1/3}\mathbf{r}^N) + \frac{\Pi^2}{2M} + pV \ , \tag{18}$$

where the first two terms on the RHS of Eq. 18 are identical to the Hamiltonian defined in Eq. 9, with the slight difference that a volume dependent scaling of the momenta, $\mathbf{p}_i$, and positions, $\mathbf{r}^N$ is introduced. The scaling makes the coordinates and the momenta of the particles explicitly volume dependent and further promotes the volume to a dynamical variable. The dynamics of the volume are controlled by an interplay between a so-called imaginary piston, which acts on the system to enforce a pressure of a desired value and the work of the system against this piston.[114] The piston has a mass M with a corresponding momentum Π and the work of the system is given by pV. The goal is that the internal pressure of system equals the external one enforced by the piston. For a better understanding, the role of the piston is illustrated in figure 2.

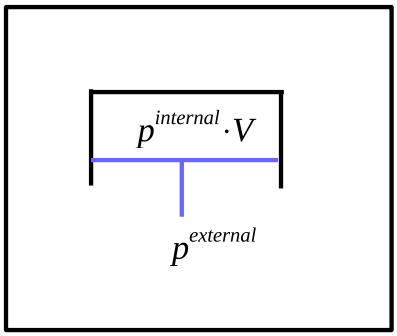


Figure 2: Illustration of the imaginary piston (blue) acting on a system to equalize internal and external pressure of a system.

The two additional DOFs - volume and momentum of the piston - change the set of Newton's equations of motion to:[114]

$$\frac{d(V^{-1/3}\mathbf{r}_i)}{dt} = \frac{\partial h_1}{\partial(V^{1/3}\mathbf{p}_i)} = \frac{V^{1/3}\mathbf{p}_i}{m_i V^{2/3}} \ , \tag{19}$$

$$\frac{d(V^{1/3}\mathbf{p}_i)}{dt} = \frac{\partial h_1}{\partial(V^{-1/3}\mathbf{r}_i)} = V^{1/3}\mathbf{f}_i(V^{1/3}\mathbf{r}_i) \ , \tag{20}$$

$$\frac{dV}{dt} = \frac{\partial h_1}{\partial \Pi} = \frac{\Pi}{M} \ , \tag{21}$$

$$\frac{d\Pi}{dt} = -\frac{\partial h_1}{\partial V} = -(p - p_V(V^{-1/3}\mathbf{r}^N, V^{1/3}\mathbf{p}_i, V)) \ , \tag{22}$$

where Eqs. 19 and 20 capture the dynamical evolution of the particles and Eqs. 21 and 22 the dynamical evolution of the volume. These equations of motions enable MD simulations at a constant pressure,

but not at constant pressure and temperature. To further maintain the temperature constant as well, Nosé-Hoover chains [119,120,123] have to be introduced to the Hamiltonian in Eq. 18: [114,124]

$$\begin{aligned} h_2(V^{-1/3}\mathbf{r}^N, V^{1/3}\mathbf{p}_i, V, \Pi) &= \sum_{i=1}^{N} \frac{V^{-1/3}\mathbf{p}_i^2}{2m_i V^{2/3}} \\ &+ U(V^{1/3}, V^{-1/3}\mathbf{r}^N) + \frac{\Pi^2}{2M} + pV + \sum_{j=1}^{M}\left[\frac{p_{\eta_j}^2}{2V_j} + \frac{p_{\zeta_j}^2}{2V_j'} + k_B T\zeta_j\right] + gk_B T\eta_1 + k_B T\sum_{j=2}^{M}\eta_j \end{aligned} \tag{23}$$

where the sum in [...] denotes the thermostatting of the volume, ζ is the volume equivalent to the external heat bath η defined in Eq. 10, which also interacts through M contributions, and the last two terms describe the thermostatting of the particles as introduced in Eq. 10. This Hamiltonian goes back to the work of Martyna, Tobias and Klein and is therefore known as the MTK barostat. [124,125] Solving Newton's equations based on this Hamiltonian leads to an exact NPT ensemble. For a more detailed explanation and derivation of the MTK barostat the interested reader is referred to the literature. [114,124,125]

2.2 Coarse Graining

A disadvantage of standard particle based MD simulations is that they are limited to system sizes of hundreds of nanometers and time scales of a couple of microseconds independent of the thermodynamic ensemble. [9] This is too small and and too short to study large scale phenomena such as self-assembly or phase transitions in soft matter systems. [2,13] To foster the study of more complex systems, one can reduce the number of DOFs available to the system. This reduction in DOFs is commonly referred to as coarse graining and should only affect those DOFs, which are considered to be irrelevant for the physical problem of interest. [25,40] In order to select the DOFs lost upon coarse graining, a so called mapping operator, $\mathbf{M}$, is applied, which groups multiple atoms to coarse grained (CG) beads or effective interaction sites, [85]

$$\mathbf{R} = \mathbf{M}(\mathbf{r}) \ . \tag{24}$$

This mapping operator maps a reference fine grained (FG) configuration, $\mathbf{r}$, onto a CG configuration space, $\mathbf{R}$, lower in resolution. Since the chosen mapping scheme sets the framework for the yet to be derived CG model, it should be selected with great care. Although this is a well-known issue, there is still no simple method available to determine an optimal scheme and the choice heavily relies on the chemical intuition of the user. [41–43] Most commonly, the position of a CG particle, I, is determined by the position of the center of mass of the FG particles, i, which are selected to belong to same CG bead. Thus, the position of the CG bead, $\mathbf{R}_I$, can be cast as [100]

$$\mathbf{R}_I = \frac{\sum_{i\in I} m_i \mathbf{r}_i}{\sum_{i\in I} m_i} \ , \tag{25}$$

where $\mathbf{r}_i$ are the position coordinates of the FG particles and m_i is the corresponding mass of the particle.

After determining the resolution of the CG model, a new set of interaction potentials has to be ascertained for the remaining degrees of freedom in the system. The exact solution to this problem would be so called multibody potential of mean force (m-PMF), whose definition in Eq. 1 is recalled [40]

$$W(\mathbf{R}) = -k_B T \ln \int_V d\mathbf{r} \exp[-\beta U_{FG}(\mathbf{r})]\delta(\mathbf{R} - \mathbf{M}(\mathbf{r})) \ , \tag{26}$$

where $W(\mathbf{R})$ is the m-PMF, k_B is the Boltzmann constant, T is the temperature of the system, V is the volume of the system, β is $1/k_BT$, U_{FG} is the potential energy of the FG system, $\mathbf{M}$ is the mapping operator and $\mathbf{r}$ and $\mathbf{R}$ are the FG and CG configurations respectively. The Dirac δ distribution acts as a filter, which accounts for the degeneracy of the CG model. In reality, the integral in Eq. 26 is too complicated to compute exactly, due to its highly multibody character. Thus, only approximate solutions are possible (see Eq. 2).[40]

There are two distinct strategies to determine an approximate solution to the m-PMF, the top-down approach[45,46] or the bottom-up approach, which is also known as systematic coarse graining.[3] The top-down approach relies on the analytical description of the effective interactions between CG particles.[51] This makes top-down models applicable for a class of systems, whose behavior can be accurately described via the analytical function chosen. The key step in top-down coarse graining is the choice of a proper analytical function and the determination of its parameters. Further, top down models also enable predictive calculations.[46] But up to some extent they lack information on chemical specificity of the underlying systems.[62,64]

On the other hand, bottom-up coarse graining allows the CG model to contain more specific information compared to the top-down approach. In bottom-up coarse graining the interactions are determined based on information obtained from the FG system, which is projected onto the CG configuration space. The class of bottom-up coarse graining methods can be further subdivided into derived methods and inverse methods.[3] The derived methods - as their name suggests - derive the effective particle interactions by projecting a mean pair force, which is free of any higher order correlations, along the distance vector between two CG beads.[80,81,126,127] The inverse methods instead generate effective interactions between two CG interaction sites in order to minimize the difference between the FG and CG model with respect to one or multiple target properties.[83,86,95–97] They explicitly contain higher order correlations, which are necessary to reproduce the target property. The importance of these higher order correlations for bottom-up coarse graining is discussed further in chapter 2.4. In the following, we define the underlying inverse problem in the context of structure based coarse graining in greater detail, as it is the main topic addressed in this thesis.

2.3 Inverse problems and structure based coarse graining

To define the inverse problem of finding a correct CG potential functional, we start with a forward problem of type

$$\mathbf{A}\vec{x} = \vec{b}\,, \tag{27}$$

where matrix $\mathbf{A}$ is multiplied with a vector $\vec{x}$ to yield a target vector $\vec{b}$. In bottom-up coarse graining, the $\mathbf{A}$-matrix, and the $\vec{b}$- vector are known and the $\vec{x}$- vector is unknown, therefore the inverse problem has to be solved.

$$\mathbf{A}^{-1}\vec{b} = \vec{x} \tag{28}$$

The optimal solution to this inverse problem, should satisfy the following consistency criteria[85]

$$P_R(\mathbf{R}) = p_R(\mathbf{R}) \tag{29}$$

which guarantees that each CG configuration $\mathbf{R}$ will be sampled with a probability P_R , which equals the mapped probability distribution $p_R(\mathbf{R})$ obtained from the underlying FG system. This mapped probability distribution is defined as[85]

$$p_R(\mathbf{R}) = z^{-1}\int d\mathbf{r}\ \exp[-\beta U_{FG}(\mathbf{r})]\delta(\mathbf{R}-\mathbf{M}(\mathbf{r}))\,, \tag{30}$$

where all variables have the same meaning as defined previously and z is the configurational integral which computes as:[85]

$$z = \int d\mathbf{r}\ \exp[-\beta U_{FG}(\mathbf{r})] \tag{31}$$

Similarly, the CG probability distribution is defined as,[85]

$$P_R(\mathbf{R}) = Z^{-1} \exp(-\beta U_{CG}(\mathbf{R}))\ , \tag{32}$$

where Z is the configurational integral over the CG configuration space.[85]

$$Z = \int d\mathbf{R}\ \exp(-\beta U_{CG}(\mathbf{R})) \tag{33}$$

In Eqs. 32 and 33, $U_{CG}(\mathbf{R})$ is the CG potential energy, which is determined by the unknown CG FF $\vec{x}$. To derive the theoretical solution to the inverse problem, we insert Eqs. 32 and 30 into Eq. 29.

$$P_R(\mathbf{R}) = Z^{-1} \exp(-\beta U_{CG}(\mathbf{R})) = z^{-1} \int d\mathbf{r}\ \exp[-\beta U_{FG}(\mathbf{r})] \delta(\mathbf{R} - \mathbf{M}(\mathbf{r})) = p_R(\mathbf{R})\ , \tag{34}$$

where the integral can be expressed in terms of the m-PMF defined in Eq. 26. This simplifies Eq. 34 to:

$$P_R(\mathbf{R}) = Z^{-1} \exp(-\beta U_{CG}(\mathbf{R})) = z^{-1} \exp[-\beta W(\mathbf{R})] = p_R(\mathbf{R}) \tag{35}$$

To obtain the CG potential energy, the inverse is taken

$$U_{CG}(\mathbf{R}) = -k_B T \ln P_R(\mathbf{R}) = -k_B T \ln p_R(\mathbf{R}) = W(\mathbf{R}) \tag{36}$$

and one sees that the consistency criteria is only fulfilled if $U_{CG}(\mathbf{R})$ equals the m-PMF, $W(\mathbf{R})$, up to an arbitrary constant. Thus, the optimal solution to the inverse problem is not accessible in practice, due to the high complexity of the m-PMF. Only approximate solutions can be obtained, which are consistent with respect to selected target properties.[85,97,98]

In the context of this thesis we focus on approximate solutions, which guarantee consistency between the CG and FG system in the pair distribution function (PDF), instead of the total configurational probability distribution.[95,97] The PDF (g_{ij}) is commonly referred to as radial distribution function (RDF) between two particles i and j. It is a measure for the ratio between the local density surrounding a central particle and the bulk density as a function of the distance between the particles. It is computed by[111]

$$g_{ij}(r) = \frac{\rho^{local}(r)}{\rho^{bulk}} = \frac{V}{N^2 4\pi r^2 \Delta r} \left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle\ , \tag{37}$$

where ρ is the density, V is the volume, N is the number of particles, r is the radius of a sphere, Δr denotes the increment by which the radius gets continuously increased, the δ function filters out particles which do not belong to the same spherical volume and r_{ij} denotes the distance between two particles. In practice, the δ function is replaced by a histogram with a specific bin size and the average is performed over a number of frames of a simulation trajectory. Now, the consistency criteria for the $g(r)$ is defined as,

$$g_{ij}^{FG}(\mathbf{M}(r)) = g_{ij}^{CG}(R) \tag{38}$$

where $g_{ij}^{FG}(\mathbf{M}(r))$ is the RDF obtained from the mapped FG configuration, $\mathbf{M}(\mathbf{r})$, and $g_{ij}^{CG}(R)$ is the RDF generated by the CG FF. As stated by Henderson, a CG FF, which fulfills this consistency criteria, is the unique solution to the inverse problem up to a constant.[128] In practice, this so-called Henderson theorem is not applicable, since several studies revealed that multiple solutions can fulfill the consistency criteria.[129–131] To determine a solution, we express the $g(r)$ as a function of the pair potential of mean force (p-PMF) $U_{ij}^0(r)$

$$g_{ij}^0(\mathbf{M}(r)) = \exp(-\beta U_{ij}^0(\mathbf{M}(r))) \ , \tag{39}$$

where β is again $1/k_BT$. By knowing $g_{ij}^0(\mathbf{M}(r))$, we can simply take the inverse

$$U_{ij}^0(\mathbf{M}(r)) = -k_BT \ln g_{ij}^0(\mathbf{M}(r)) \tag{40}$$

to determine the effective interactions in the CG configuration space. But, this simple approach will not fulfill the consistency criteria defined in Eq. 38, since the $g(r)$ is not only determined by simple pairwise interactions, but also includes higher order correlations, which have to be mapped onto the effective pair interaction. In order to improve the solution, a standard Newton inversion scheme can be applied:

$$x^n = x^{n-1} + f'(x^{n-1})\Delta x \tag{41}$$
$$\Delta x = x^0 - x^{n-1} \tag{42}$$

Based on an initial guess, x^0, the unknown x-value gets updated a series of n-times following a first order Taylor approximation until the consistency criteria is fulfilled. In Eq. 41, f' is the first order derivative of a function $f(x)$ with respect to x.
The two most common structure based coarse graining methods, which use such an iterative scheme are the iterative Boltzmann inversion (IBI)[95,96] and the inverse Monte Carlo (IMC)[97] approach.

In IBI, the p-PMF gets updated according to:

$$U_{ij}^n(R) = U_{ij}^{n-1}(R) + \Delta U_{ij}(R) \ , \tag{43}$$
$$\Delta U_{ij}(R) = U_{ij}^0(\mathbf{M}(r)) - U_{ij}^{n-1}(R) = k_BT \ln \left[\frac{g_{ij}^{n-1}(R)}{g_{ij}^0(\mathbf{M}(r))} \right] \tag{44}$$

This approach is an approximate Newton method to solve the inverse problem, since $f'(U^{n-1})$ is assumed to be the identity matrix. The correct first order derivative for an exact Newton scheme would be:[132]

$$f'(U^{n-1}) = \frac{\partial \ln g_{ij}}{\partial U_{ij}} = \frac{1}{g_{ij}} \frac{\partial g_{ij}}{\partial U_{ij}} = -\frac{1}{k_BT} = -\beta \cdot Id \ , \tag{45}$$

which is only exact in the low density limit. Thus, IBI gives approximate solutions to the inverse problem.

Contrary to that, the IMC method of Lyubartsev and Laaksonen[97] provides an exact Newton inversion method to solve the inverse problem of finding a structural consistent CG FF. In IMC a discretized Hamiltonian is assumed[97]

$$H(q) = \sum_\alpha U_\alpha N_\alpha(q) \ , \tag{46}$$

where q are the particle coordinates, U_α is the effective pair potential and N_α is the corresponding number of particle pairs at a specific distance α. The number of particle pairs is related to the $g(r)$ via:[97]

$$N_\alpha = \frac{N(N-1)}{2}\frac{4\pi r^2 \Delta r}{V} g(r) \tag{47}$$

where all variables have the same meaning as defined in Eq. 37. In IMC, the iterative update on the p-PMF computes as follows:[97]

$$U_{ij}^n(R) = U_{ij}^{n-1}(R) + f'(U_{ij}^{n-1}(R))\Delta U_{ij}(R)\ , \tag{48}$$

where the second term on the RHS is given by:

$$f'(U_{ij}^{n-1}(R))\Delta U_{ij}(R) = \frac{\partial \langle N_\alpha^{CG}\rangle}{\partial U_\gamma}(U_{ij}^0(\mathbf{M}(r)) - U_{ij}^{n-1}(R)). \tag{49}$$

In IMC, the average number of particle pairs at a distance α is computed as[97]

$$\langle N_\alpha \rangle = \frac{\int dq N_\alpha(q) \prod_\lambda \exp[-\beta U_\lambda N_\lambda(q)]}{\int dq \prod_\lambda \exp[-\beta U_\lambda N_\lambda(q)]}\ . \tag{50}$$

This enables the formulation of a first order derivative, which is not only valid in the low density limit:[97]

$$\frac{\partial N_\alpha}{\partial U_\gamma} = -\beta \cdot (\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \langle N_\gamma \rangle) \tag{51}$$

Thus, IMC is an exact Newton inversion method to determine structural consistent CG models.
A study on practical implications caused by the differences between IBI and IMC is presented in chapter 3.1[131] of this thesis.

2.4 Representability and transferability of coarse grained models

Independent of the applied coarse graining approach, all CG models try to simultaneously achieve accurate representability and transferability. This means, the derived models should quantitatively reproduce the underlying FG model at a given state point and they should be transferable to different state points in the thermodynamic and chemical phase space as well.[82]
The difficulty to achieve both arises from the embedding of entropic contributions into the effective pair interactions. The m-PMF and its corresponding approximate solutions can be split into an energetic and entropic contribution, as any other free energy.[133]

$$W(R) = U_W(\mathbf{R}) - TS_W(\mathbf{R}) \tag{52}$$

To achieve accurate representability of the FG configuration space at a given state point, both contributions have to be effectively captured by the generated CG pair interactions. When going to a different state point, especially the entropic part can change significantly. The effective pair potentials have to be able to account for this change, otherwise the CG model is not transferable. Thus, getting the right balance in the entropic contributions to the PMF is key to simultaneously achieve representability and transferability.[40]

Moreover, by determining a CG potential via the inverse route, quantitative representability is only guaranteed for the specific target property chosen.[134] This is owed to the mapping of all higher order correlations, which contribute to this specific property, onto a simple pairwise additive potential.[102,103] This means, reproducing the FG structure does not guarantee a match in thermodynamic properties, such as pressure,[96] surface tension[135] or Kirkwood-Buff integrals,[110] or dynamic properties, such as the diffusion coefficient.[104] Correspondingly, a match in the pair forces does not automatically lead to a match in the pair structure.[80]

To improve this representability problem, the use of additional thermodynamic constraints has been proven to be beneficial. In many cases a linear ramp potential, ΔU_{Ramp}, has been added to the derived effective pair potential in order to reproduce a second target property X as well.[96,104,110] ΔU_{Ramp} is usually calculated based on the difference between X in the FG (X^{FG}) and CG system (X^{CG}).

$$\Delta U_{Ramp} = \alpha'(X^{CG} - X^{FG})\left(1 - \frac{r}{r_{cut}}\right) \tag{53}$$

In Eq. 53 α' is an empirical prefactor, r is the separation distance between particles and r_{cut} is equivalent to the cut-off of the effective pair potential. Ramp corrections have been successfully tested to improve the representability of the virial pressure[96] or Kirkwood-Buff integrals,[110] without losing accuracy in the structural representability. Despite their practical utility, these constraints are added to the effective pair potentials after they got generated. This post-processing formally violates the consistency criteria defined in the previous section (see. Eqs. 29 and 38). To circumvent this problem, one can directly constrain the coarse graining algorithm itself, either by the use of Lagrange multipliers[135,136] or via Gaussian elimination techniques.[137] Here, a solution is derived under the constraint of reproducing a second target property implicitly. Constraint algorithms have been successfully tested for reproducing the virial pressure[137] and/or the surface tension[135,136] in the context of different coarse graining methods.

Instead of constraining the effective pair potentials,[138] one could also extend the system's Hamiltonian in order to improve the representability of a CG model. The idea of extended Hamiltonians has first been proposed by Ashcroft and Stout in the context of liquid metals,[139] and has later been adapted in the field of soft matter physics.[1,105,140–142] The basic philosophy is the addition of supplementary potential energy terms to compensate for the loss of DOFs upon coarse graining,

$$H(\mathbf{R}^N, \mathbf{p}^N, X) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{R})^N + U_X(X)\ . \tag{54}$$

In Eq. 54 the first two terms on the RHS correspond to the kinetic and potential energy of the system, as previously defined in Eq. 9, the last term corresponds to an additional energy contribution, which for example can depend on the system's volume[143] or a local density[144] as further discussed in chapters 3.2, 3.3 and 3.4 of this thesis.

All these approaches should not only improve the representability, but also the transferability of the derived CG models. Ramp corrections for the Kirkwood-Buff integrals, for example improved the concentration transferability.[110] Extended Hamiltonian approaches could improve the transferability to different concentrations[145] and temperatures[146] or from bulk to interfacial systems.[147]

Another possibility is the use of an extended ensemble parametrization, instead of an extended Hamiltonian approach. Here, the derived effective p-PMF is obtained, by averaging the solutions of the inverse problem defined at different state points.[108]

$$\overline{U}_\Gamma(\mathbf{R}_\Gamma) = \sum_\zeta \sum_{\lambda_\Gamma} U_\zeta(\psi_\zeta(\{\mathbf{R}\}_{\lambda_\Gamma}))\ , \tag{55}$$

where $\overline{U}_\Gamma(\mathbf{R}_\Gamma)$ is the averaged effective p-PMF, Γ defines the different state points, $\mathbf{R}$ is the CG configuration space, ζ describes the type of interaction, e.g. bond or non-bonded interaction, U_ζ is the solution of the inverse problem at a given topology and for a given interaction type, ψ_ζ is the corresponding scalar variable, e.g. bond or pair distance, and λ_Γ are the set of CG sites existing at a given topology.
Extended ensemble approaches have been successfully tested for structure-based[109] and force-based coarse graining methods,[148] which in the latter example resulted in a concentration transferable and in the first case in a temperature/density transferable CG model.
For completeness, there are also methods which try to tackle the representabilty/transferability problem by developing effective interactions for different state points.[149] These potentials are then combined either via a simple interpolation scheme[150] or a switching function, which based on some internal state weighs those potentials differently.[151–153]
Despite the success of all these approaches to improve either the representability or the transferability or both, there is still no general solution to the representability/transferability problem.[154] This issue is going to be further discussed in the following results section.

Results

In this section the main results of this thesis are presented. The four peer-reviewed articles are ordered based on their topics and not on their chronological appearance.

The first article presents a methodological comparison between IBI-type methods and IMC. We investigate a binary mixture of particles different in size, which only interact via a 6-12 Lennard-Jones (LJ) potential. This study contains a deep analysis on theoretical and numerical differences between the different methods and the effect of these differences on the convergence of the methods and on the representability of the derived models. On the basis of the findings of this work, IMC is the method chosen for all remaining projects. The article is reprinted by permission from Springer Nature: The European Physical Journal - Special Topics: Comparison of iterative inverse coarse-graining methods, David Rosenberger, Martin Hanke, Nico F. A. van der Vegt, 2016, 225, 8, 10 October 2016 (doi:10.1140/epjst/e2016-60120-1).

The following work shifts the focus away from the representability towards the concentration transferability. Again we investigate the binary mixture of LJ particles, but now at different mole fractions of the larger component. Here, we show that it is possible to achieve concentration transferability for IMC based models, if the model is derived at an ideal state point. This state point can be identified with the relative entropy.[98] In addition, we show that trends in transferability observed for a model system, can also be seen for an IMC model for binary mixtures of hexane and perfluorohexane. This work has been just accepted for publication in Phys. Rev. E (May 6th, 2019).

In the third chapter we investigate the effect of local density dependent potentials on the concentration transferability of CG models for methanol in implicit water and water in implicit methanol. Here, we show that including local density potentials only improves the transferability of the CG model if the water is explicit and only in the direction of increasing water concentration. Thus, changes in local environment play a less dominant role on changes in the liquid structure of mixtures of water and methanol, if the methanol concentration gets increased. This illustrates the power of bottom-up coarse graining to identify negligible effects on the system's driving force. Additionally, we show that IMC and relative entropy optimization yield the same set of CG potentials up to numerical uncertainties.

In the last chapter a methodological Ansatz is presented, which enables the development of temperature transferable IMC models for liquid alkanes. Instead of using a local density dependent potential, here a volume dependent potential is added to the IMC model. We propose a new approach to parametrize this volume dependent potential, which leaves the IMC model untouched and is based on simple linear regression. This method yields IMC models which are transferable almost over the whole temperature range of the liquid phase, from the melting point towards the evaporation temperature. This has not been achieved before with only one set of effective pair potentials. The corresponding article is reproduced from Ref. [145] with permission from the PCCP Owner Societies.

Eur. Phys. J. Special Topics

DOI: 10.1140/epjst/e2016-60120-1

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Comparison of iterative inverse coarse-graining methods

David Rosenberger[1], Martin Hanke[2] and Nico F.A. van der Vegt[1,a]

[1] Eduard-Zintl-Institut für Anorganische und Physikalische Chemie and Center of Smart Interfaces, Technische Universität Darmstadt, Alarich-Weiss-Strasse 10, 64287 Darmstadt, Germany
[2] Institut für Mathematik, Johannes Gutenberg Universität Mainz, 55099 Mainz, Germany

Received 15 April 2016 / Received in final form 25 June 2016
Published online XX XXX 2016

Abstract. Deriving potentials for coarse-grained Molecular Dynamics (MD) simulations is frequently done by solving an inverse problem. Methods like Iterative-Boltzmann-Inversion (IBI) or Inverse Monte-Carlo (IMC) have been widely used to solve this problem. The solution obtained by application of these methods guarantees a match in the radial distribution function (RDF) between the underlying fine-grained system and the derived coarse-grained system. However, these methods often fail in reproducing thermodynamic properties. To overcome this deficiency, additional thermodynamic constraints such as pressure or Kirkwood-Buff-Integrals (KBI) may be added to these methods. In this communication we test the ability of these methods to converge to a known solution of the inverse problem. With this goal in mind we have studied a binary mixture of two simple Lennard-Jones (LJ) fluids, in which no actual coarse-graining is performed. We will further discuss whether full convergence is actually needed to achieve thermodynamic representability.

1 Introduction

Molecular Dynamics (MD) simulations can be used to follow the microscopic dynamics of atomic-scale processes up to several hundreds of nanoseconds. While this may be sufficient for many condensed phase systems, many properties of soft condensed matter are determined by processes on time and length scales much longer than that [1–5]. Owing to high computational cost of sampling a vast number of atomistic degrees of freedom, the relevant time and length scales in soft matter are often inaccessible with classical, atomistic MD models. To make these scales accessible with current computational power the number of degrees of freedom (DOF) of the models have to be reduced to a degree where a correct physical description of the system is still guaranteed. This process is commonly referred to as coarse-graining.

Coarse-graining can be achieved with bottom-up procedures in which degrees of freedom of an detailed-atomistic, or fine-grained (FG), system are systematically

[a] e-mail: vandervegt@csi.tu-darmstadt.de

integrated out [6,7]. A major challenge in systematic coarse-graining is the representability of the coarse-grained (CG) model, i.e. observables computed with the CG model and its parent FG model must be the same. In practice, however, full representability cannot be achieved because the multibody potential of mean force of the CG degrees of freedom is usually approximated by effective pair potentials [8–16].

The focus of this work is on iterative inverse coarse-graining methods. Methods of this class derive effective pair potentials which aim to reproduce certain target properties of the FG system. Common target properties are the radial distribution functions (RDFs). The derivation of an effective pair potential reproducing the RDF constitutes an inverse problem. According to the Henderson theorem [17] this inverse problem has a unique solution, and the two most common methods to solve this problem are the Iterative Boltzmann Inversion (IBI) [11] and Inverse Monte Carlo [12] (IMC) methods. Both methods use an iterative scheme in which an initial guess for the effective pair potential is updated until convergence between the RDF of the CG system and the RDF of the FG system is achieved. The representability of the CG model however crucially depends on how well the effective pair potentials approximate the multibody potential of mean force associated with the coarse variables and has been discussed by Noid et al. [18,19] and others [20–23]. While some approaches use thermodynamic constraints, selected to represent quantities like the pressure in the iterative optimization procedure [11,24–26], other coarse graining approaches have emphasized properties of the pair potentials in terms of the underlying, effective, atomistic interactions [21,27,28]. While application of thermodynamic constraints in iterative inversion methods may interfere with the Henderson theorem, in practice it is often considered more important to recover thermodynamic quantities than to satisfy the Henderson theorem. IBI-derived models have been derived over the last years for various soft matter systems, including polymers [11,29–31] and liquid crystals [32].

A systematic study of the method itself was previously reported by Jain et al. and Fu et al. [33,34] They tested the capability of IBI to converge to a known solution for the inverse problem. In the two studies, a simple one-component Lennard Jones (LJ) fluid was taken as a reference or FG system. Both investigated the possibility to recover the potentials that were applied to generate the target RDFs without loosing any DOF. The study of Jain et al. shows that even after 1000 iterations full convergence in the potential has not been accomplished, whereas the structures have converged already after 10 iterations. This is due to the fact that the structure of such a simple fluid is mainly determined by short ranged repulsive interactions. This part of the interaction potential converges much faster than the attractive long range part. To account for the attractive part constraints can be added to the standard IBI scheme. Application of these constraints leads to a faster convergence towards the correct potential [33]. Fu et al. extended the study [34]. They tested the influence of the cut-off and the capability of the effective pair potential to reproduce the isothermal compressibility. The authors suggest to choose the cut-off such that the oscillating long range part of the RDF is removed, since the method is insensitive to this part. Moreover, it can be shown that effective pair potentials based on IBI lead to lower isothermal compressibility compared to the FG system. With an additional constraint added to the method the compressibility increases compared to the FG system, but is closer to the target value [34].

In this paper we extend these studies for a binary mixture of LJ fluids without loosing any DOF. We systematically show the strengths and weaknesses of the IBI method with and without additional constraints and compare them with the IMC method. We furthermore want to address to what degree convergence is needed to achieve appropriate thermodynamic representability with the generated effective pair potentials.

2 Methods

2.1 Iterative Boltzmann inversion

Generation of effective pair potentials for non-bonded interactions which guarantee structural agreement at a pair level between the FG and CG system can be achieved by application of the IBI technique [11]. Here, an effective pair potential is derived from the potential of mean force (PMF)

$$U_{ij}^{0}(r) = -k_B T \ln g_{ij}^{ref}(r) \tag{1}$$

between molecules of two species i and j separated by the distance r, where g_{ij}^{ref} is the reference RDF between the two species i and j, k_B is the Boltzmann constant and T is the temperature of the system. The initial guess is usually a poor choice and lacks the ability to reproduce the target RDF. For that U_{ij}^{0} gets updated (see Eq. (2)) n-times until convergence between the target RDF and the RDF generated from the effective pair potential is reached.

$$U_{ij}^{n}(r) = U_{ij}^{n-1}(r) + k_B T \ln \left[\frac{g_{ij}^{n-1}(r)}{g_{ij}^{ref}(r)} \right] = U_{ij}^{n-1}(r) + \Delta U_{ij}(r) \tag{2}$$

Recent work by Ivanizki revealed that IBI is a modified Newton method [35] based on an approximation of the Jacobian

$$\frac{\partial \log g_{ij}}{\partial U_{ij}} = \frac{1}{g_{ij}} \frac{\partial g_{ij}}{\partial U_{ij}} \approx -\frac{1}{k_B T} I, \tag{3}$$

I denoting the identity matrix, which becomes exact in the low density limit where $g_{ij} = \exp(-U_{ij}/k_B T)$.

RDFs are predominately determined by short ranged repulsive interactions, i.e. a match in the RDF guarantees a match in the repulsive part of the potential between the FG and CG system. On the other hand, thermodynamic properties like pressure are mainly determined by long range attractive interactions, which are not well represented by IBI potentials. As a result, IBI potentials often overestimate the virial pressure of the system.

It should be emphasized that, in general, the incorporation of thermodynamical constraints violates the Henderson theorem, because there is only one effective pair potential that reproduces the RDF, and the thermodynamical constraints are either satisfied automatically, or they simply do not hold true for the Henderson potential [17]. Therefore, thermodynamical constraints can only be incorporated by relaxing the accuracy with which the RDF is being matched.

For IBI such a relaxation is difficult to realize because the whole iteration process is designed to target for a precise match of the RDF. Accordingly, it has been suggested to intertwine the IBI iteration with somewhat artificial correction steps to enforce the validity of thermodynamical constraints. For example, to account for the correct pressure and to improve on the attractive interactions, an additional update

$$\Delta V_{ij}(r) = \alpha \left(1 - \frac{r}{r_{cut}} \right) \tag{4}$$

can be added to the standard IBI procedure [11]. This correction is a simple linear ramp correction that becomes 0 at the cut-off r_{cut} and α at $r = 0$, where the exact form of α is discussed in Sect. 3.2. From now this method will be referred to as P-IBI.

Besides this simple linear polynomial there are more complex functions that can be used for ΔV_{ij} as well. These functions take the density of the system into account or try to ensure correct pressure via a variational principle [24–26]. Inclusion of the pressure correction changes Eq. (2) to:

$$U_{ij}^{n}(r) = U_{ij}^{n-1}(r) + \Delta U_{ij}(r) + \Delta V_{ij}(r). \tag{5}$$

It should be pointed out that every gain comes with a cost. The cost in the case of having matched pressure as well as the RDF is the loss of exact representation of the compressibility at a coarse-grained resolution [24].

Besides the RDF another property can be introduced as a target property namely the Kirkwood-Buff integrals (KBI). These integrals can be related to a variety of other thermodynamic properties of multicomponent systems like activity coefficients, solvation free energies, partial molar volumes and the isothermal compressibility [36]. KBIs are integrals of the RDF over a finite volume [37,38], namely

$$G_{ij}^{V} = \frac{1}{V}\int_{V}\int_{V}[g_{ij}^{\mu VT}(r_{12}) - 1]d\boldsymbol{r}_1 d\boldsymbol{r}_2\,. \tag{6}$$

In Eq. (6), $g_{ij}^{\mu VT}$ is the RDF in the grand-canonical ensemble. In the limit $V \to \infty$, the integral in Eq. (6) changes to [38]

$$G_{ij}^{\infty} = 4\pi\int_{0}^{\infty}[g_{ij}(r) - 1]r^2 dr \tag{7}$$

where G_{ij}^{∞} is the KBI in the thermodynamic limit. The KBIs (Eq. (7)) should be formally calculated in grand-canonical simulations, but it was shown by several research groups that the calculation of KBIs is also possible under NpT or NVT conditions [39–48]. To this end, the running Kirkwood-Buff integral (RKBI), defined as

$$G_{ij}^{V}(r) = 4\pi\int_{0}^{r}[g_{ij}(s) - 1]s^2 ds, \tag{8}$$

is calculated and G_{ij}^{∞} is approximated by averaging the RKBI in a suitable range of r, typically between 1.0 nm and 2.0 nm in aqueous solution mixtures, where the RKBI reaches a limiting plateau value. The justification for this lies first within that density fluctuations are only local far away from the critical point and second within that correlation distances between the particles are small enough compared to the box size [48].

The idea proposed by Ganguly et al. is to use RKBIs within the framework of IBI. This method is called Kirkwood-Buff Iterative Boltzmann Inversion (KB-IBI) [47]. The main objective of this method is to guarantee not only structural, but also thermodynamic representability at a coarse-grained level, for systems with multiple components. In case of KB-IBI the update is performed according to a ramp correction as [47]

$$\Delta KBI_{ij}(r) = \alpha'(G_{ij}^{n-1} - G_{ij}^{ref})\left(1 - \frac{r}{r_{cut}}\right) \tag{9}$$

where G_{ij}^{ref} is the plateau value of the RKBI calculated from the reference simulations and G_{ij}^{n-1} is the plateau value of the RKBI after the $n-1$th step. The values for G_{ij} are obtained by averaging the RKBI in a range where the oscillations around its plateau value are small. The parameter α' is a system specific prefactor, which, as a good estimation, lies in the range between 0.01 and 0.1 kJ nm^{-3} mol^{-1} [47].

The KB-IBI ramp correction (Eq. (9)) can be combined with the regular IBI update $\Delta U_{ij}(r)$, in a way similar to the pressure ramp correction in Eq. (5).

Finally, the running integral of the RDF can be considered to compute the update

$$\Delta C_{ij}(r) = k_B T \ln \left[\frac{\int_0^r g_{ij}^{n-1}(s) s^2 ds}{\int_0^r g_{ij}^{ref}(s) s^2 ds} \right]. \tag{10}$$

This method is called C-IBI according to Oliveira et al. [49] The potential gets updated n times until convergence between the running integrals of the FG and CG system is reached.

$$U_{ij}^n(r) = U_{ij}^{n-1}(r) + \Delta C_{ij}(r). \tag{11}$$

2.2 Inverse monte carlo

The IMC method developed by Lyubartsev and Laaksonen provides another way to derive effective pair potentials which reproduce a target RDF [12]. In contrast to IBI, IMC is an exact Newton method [50] and aims to directly fit g_{ij} rather than $\log g_{ij}$ as compared to Eq. (2), i.e.,

$$U_{ij}^n = U_{ij}^{n-1} - A^{-1}(g_{ij}^{n-1} - g_{ij}^{ref}), \tag{12}$$

where

$$A = \frac{\partial g_{ij}}{\partial U_{ij}} \tag{13}$$

is the corresponding Jacobi matrix. As shown in [12,51,52] this matrix consists of cross correlation quantities; it can therefore be assembled during the forward simulation, but this requires longer simulation times to achieve sufficiently good statistics.

Instead of solving Eq. (12) exactly, one can reformulate this as a variational problem and add some penalty term to enhance stability [53]:

$$\Delta U_{ij} = \arg\min \|A\Delta U_{ij} - (g_{ij}^{n-1} - g_{ij}^{ref})\|^2 + \lambda \|R\Delta U_{ij}\|^2, \tag{14}$$

because A is often very close to being singular. Eq. (14) is known as Tikhonov regularization [54]; $\lambda > 0$ is the regularization parameter and R is the regularization operator, which, for example, can be the identity matrix. With this choice of R Eq. (14) can be rewritten as a linear system

$$(A^T A + \lambda I)\Delta U_{ij} = A^T (g_{ij}^{n-1} - g_{ij}^{ref}). \tag{15}$$

As mentioned before, the major advantage of IMC compared to IBI is the exact update scheme, in which the elements of the Jacobian matrix can be expressed in terms of fluctuation quantities following from statistical thermodynamics, whereas IBI uses an empirical update scheme. Another difference is that for a multi-component system every individual IMC potential update depends on the match of all target RDFs simultaneously. This means that in case of a mismatch of just one of the RDFs, all interaction potentials that are present in the system are affected. In IBI this interdependence is not given for multi-component systems. Finally, thermodynamic constraints are fairly easy to incorporate into the IMC scheme by augmenting the scheme in Eq. (14) to the effect that the minimization process is restricted over ΔU_{ij} candidates that maintain a preassigned thermodynamic quantity, as it was previously done for the surface tension [53]. To enforce the correct virial pressure of a single

Table 1. Interaction parameters for the reference binary Lennard-Jones system.

Interaction	σ [nm]	ϵ [kJ/mol]
LJ1–LJ1	0.340	0.98
LJ2–LJ2	0.296	0.84
LJ1–LJ2	0.317	0.91

component molecular fluid one could restrict the potential update ΔU of Eq. (14) to satisfy the orthogonality constraint

$$\int_0^\infty \Delta U'(r) r g^{ref}(r) r^2 dr = 0. \tag{16}$$

But these advantages come with some disadvantages. IMC needs more statistics to lead to reasonable results and therefore it requires higher computational cost per iteration. Furthermore, regions of poor sampling might perturb the stability of the algorithm and thus have to be removed [55].

3 Simulation setup

All simulations were performed with Gromacs 4.6.5 [56]. For the generation of the effective pair potentials the Versatile Object-Oriented Toolkit for Coarse-Graining Applications (VOTCA version 1.3_rc1) [55,57,58] with Gromacs 4.6.5 [56] was applied.

3.1 Reference simulation

The target RDFs were calculated from a binary mixture of Lennard-Jones particles (interaction parameters see Table 1). There were 1000 particles per type in a cubic box with an initial side length of 5 nm. The simulation time was 50 ns with a time step of 2 fs. For integration of the equations of motion the stochastic dynamics (sd) integrator of Gromacs 4.6.5 was used with a time constant of 0.5 ps^{-1} [56]. LJ interactions were treated with the cut-off method and were truncated at a distance of 1.2 nm. The simulation was run in a NpT ensemble at 1 bar and 85 K with the Parrinello-Rahman barostat ($\tau_p = 0.5$ ps) [59,60]. Periodic boundary conditions were applied in x, y and z direction. The analysis of the RDFs was performed over the last 25 ns.

3.2 Generation of effective pair potentials for Non-Bonded interactions via iterative inverse methods

The target RDFs were Boltzmann inverted and used as an input for the methods described in Sect. 2. It is crucial at this point to remind again that in the study presented here no real coarse-graining is performed. The major objective is to test the ability of the methods to reproduce the generating potentials of the target RDFs. All simulations were run under NVT conditions taking the average box size from the reference simulation and its final configuration as the starting configuration. IBI simulations without any additional constraints were run for 1 ns per iteration for 300 iteration steps. The time step for the integration was 2 fs. As an integrator the sd integrator of Gromacs 4.6.5 was used [56]. There were three different cut-off values tested for the LJ interactions: 1.2 nm, 0.85 nm and 0.60 nm. All three interactions

Table 2. Interval for RDF/interaction evaluation in the IMC scheme.

Interaction	r(0) [nm]	r_{cut} [nm]
LJ1–LJ1	0.32	0.85
LJ2–LJ2	0.27	0.85
LJ1–LJ2	0.29	0.85

were updated according to Eq. (2) in each iteration step. The same setup was used for P-IBI. The pressure was corrected by application of Eq. (4) every third iteration, where an additional scaling factor, f, of 0.0003 is included in the pre-factor α (see Eq. (4)). The pressure correction is applied in VOTCA version 1.3 according to [61]

$$\alpha = -\text{sign}(\Delta p) 0.1 k_B T \min(1, f|\Delta p|) \tag{17}$$

where Δp is the pressure difference between the actual value and the target value. For the KB-IBI update we followed the recipe of Ganguly et. al. presented in the supporting information of [47]. 30 preliminary iterations of P-IBI were followed by 50 iterations of KB-IBI update according to Eq. (9). The scaling factor α' was gradually decreased from 0.05 for the first 10 iterations to 0.03 for the next ten iterations and finally to 0.01 for the final 30 iterations. RKBIs were averaged between 0.5 nm and 0.85 nm. During the C-IBI steps the potential was updated according to Eq. (11). C-IBI was run for 100 steps and the integrals were evaluated between 0.2 and 0.85 nm.

The IMC simulations were performed without having any thermodynamic constraint included. The run time per iteration has been increased up to 5 ns. To guarantee stability of the IMC scheme regions of insufficient sampling were removed from each RDF (interval for RDF/interaction evaluation see table 2). Parameter like timestep and integrator were kept the same as in the IBI steps. The potential updates were performed in cycles, i.e only one interaction got updated per step. Further, we tested the influence of 30 preliminary steps of P-IBI followed by 4 IMC updates per interaction. The update scheme was additionally stabilized by the application of Eq. (14) with regularization parameter λ of different magnitudes (100, 300, 1000).

3.3 Calculation of thermodynamic properties

The generated effective pair potentials have been applied to calculate the following thermodynamic properties: pressure, KBIs, isothermal compressibility, the partial molar volume and the derivative of the activity coefficient. To this end, 100 ns of MD simulations were performed in the NVT ensemble. For integration of the equations of motion the sd integrator was applied [56]. The time step for the integration was 2 fs. The reference temperature was 85 K. The cut-off for the LJ interactions was chosen according to the cut-off within the generation procedure. This means for the derived potentials a cut-off of 0.85 nm was used. The analysis of the data has been performed over the last 90 ns.

4 Results and discussion

In Fig. 1 the target RDFs for the three interactions (dashed lines) present in the system are shown together with their underlying LJ potentials (solid lines, U_{ref}).

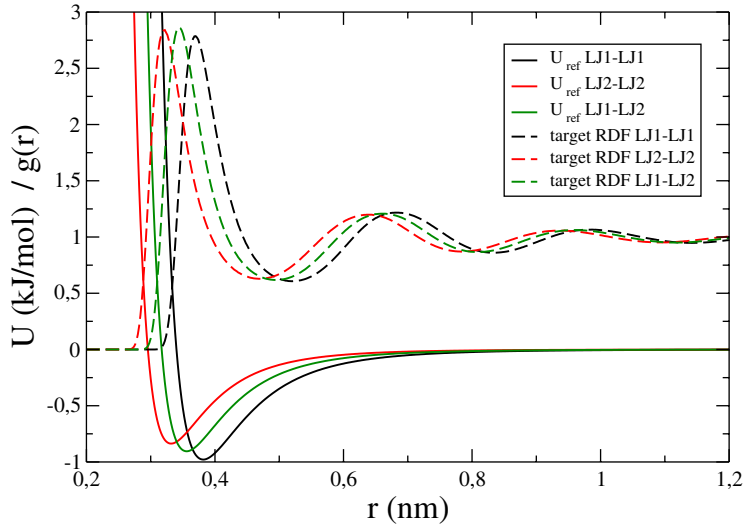


Fig. 1. Reference LJ potentials (solid lines) and target RDFs (dashed lines) for the three interactions present in the system: LJ1-LJ1 (black), LJ2-LJ2 (red) and LJ1-LJ2 (green).

4.1 Effective pair potentials via different variations of the Iterative Boltzmann Inversion method

The target RDFs were Boltzmann inverted and the different variations of the IBI method were executed as described in Sect. 3.2. For convenience the results are discussed based on the LJ1-LJ1 interaction. Since the remaining interactions in the system (LJ2-LJ2, LJ1-LJ2) show similar behaviour, the following discussion can be assigned to them accordingly.

In Figs 2 and 3 the influence of the pressure correction on the convergence of the IBI method with a cut-off of 0.85 nm is shown. One can see clearly that convergence in the RDF has been reached after 50 steps for IBI (Fig. 2) as well as for P-IBI (Fig. 3). But in both cases the potential has not converged to the reference potential. Without pressure correction one obtains less attractive potentials after the IBI process in comparison to the reference potential. This is in agreement with the fact that RDFs are mainly determined by the repulsive part of the LJ potential. By approaching a match between the RDFs the short ranged repulsive part of the potential converges faster than the long ranged attractive part. As a consequence the virial pressure of the system is overestimated by the IBI potentials. As one can see further, even after 300 steps the potential well is not matched, although an agreement in the RDF has been accomplished after 50 steps. Further iterations would lead to a closer approximation towards the reference potential, but full convergence cannot be reached within a reasonable number of iterations.

To account for the large virial pressure and the lack of attraction in the resulting potential the linear pressure correction (see Eq. (4)) is applied. This correction in P-IBI leads to a final potential which is slightly more attractive than the reference potential, but it is in better agreement with it as the potential obtained from IBI. Moreover, the method seems to have converged to a stable solution after 80 steps, i.e. further iteration does not significantly improve the obtained potentials. What we have shown up to this point is in accordance to what is known from literature for single component LJ systems [11,33,34]. Since IBI treats each interaction independently this is not a surprise. Independent treatment means that an update of one interaction does not affect the other interactions in the system [55].

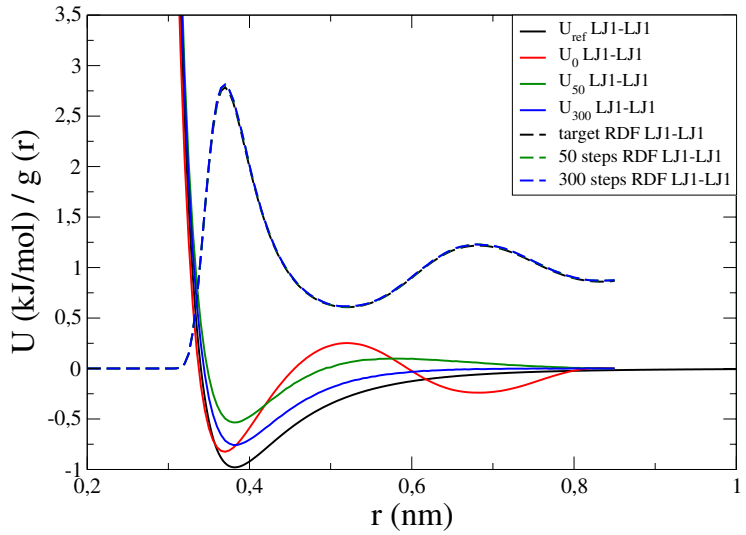


Fig. 2. IBI: RDFs (dashed lines) and potentials (solid lines) for the LJ1-LJ1 interaction: reference or target (black), the PMF (red), after 50 iterations (green) after 300 iterations (blue)

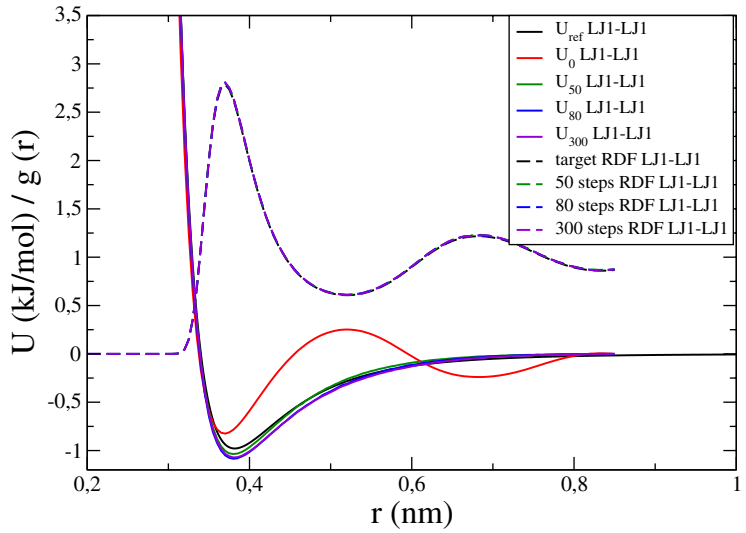


Fig. 3. P-IBI: RDFs (dashed lines) and potentials (solid lines) for the LJ1-LJ1 interaction: reference or target (black), the PMF (red), after 50 steps (green), after 80 steps (blue), after 300 steps (violet)

In the next step we investigate the influence of different cut-off values on the convergence of P-IBI. The results are shown in Fig. 4. In all three cases the RDF is well-matched after 50 iterations. If the cut-off is short (0.60 nm), the depth of the potential well closely matches with the reference potential, but the potential decays too rapidly beyond the minimum. If the cut-off is chosen to be as long as in the reference simulations (1.20 nm) a secondary minimum has evolved. As mentioned before, RDFs are mainly determined by short range repulsive interactions. Therefore,

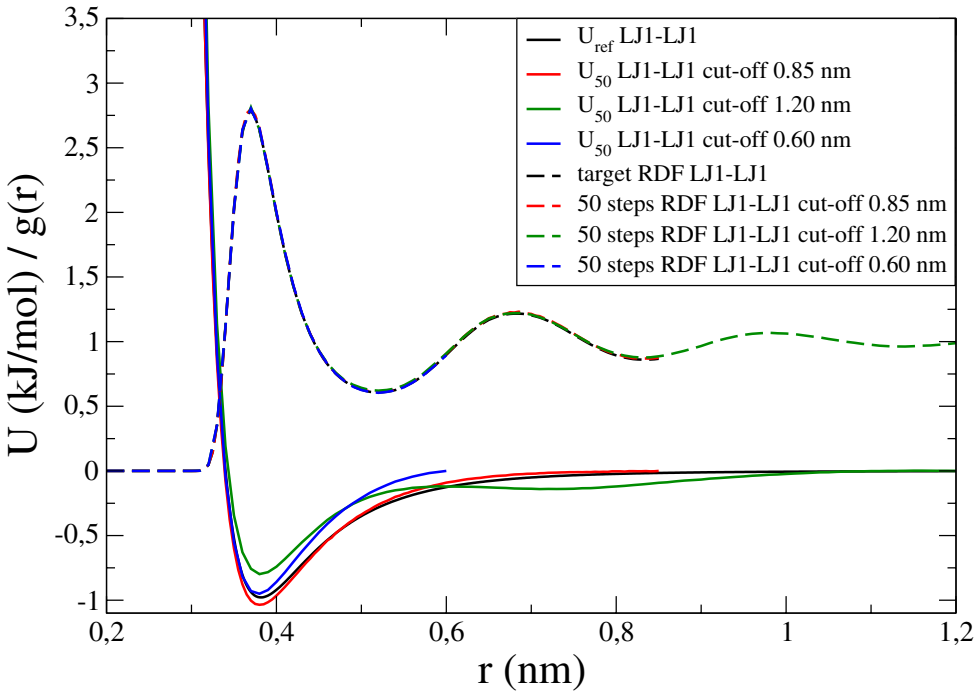


Fig. 4. Influence of different cut offs for P-IBI for the LJ1-LJ1 interaction: potentials (solid lines) and RDFs (dashed lines) target or reference (black), cut-off 0.85 nm (red), cut-off 1.20 nm (green), cut-off 0.6 nm (blue)

inclusion of long range correlations in IBI iterations leads to a poor convergence in that part of the potential and this is exactly what we observe for a cut-off of 1.20 nm. If the cut-off is chosen at the intermediate value (0.85 nm), the final potential is a bit more attractive compared to the shorter cut-off, but converges faster in the long range part compared to the longer cut-off.

To determine the optimal cut-off distance for the LJ interaction throughout the IBI process, one should choose it long enough to get good representation of the potential beyond the minimum, but also short enough to exclude fluctuations on the very long distances, which lead to poor convergence of the method. Again this is in accordance to what is known from literature for a single component system [34].

In the last step, KB-IBI and C-IBI updates have been investigated and compared to the previous results. As we see in Fig. 5, the RDF gets well reproduced by all IBI methods tested. By having a look at Fig. 6 one sees that P-IBI and IBI lead to larger RKBIs compared to the FG reference. The blue (IBI) and the violet (P-IBI) RKBIs are always above the black target RKBI. KB-IBI updates do not provide a significant improvement (red curve). If the update gets performed according to the C-IBI method, the RKBI is matched the best, especially in the longer range ($r > 0.60$ nm), but in this case the potential (solid green line in Fig. 5) becomes too repulsive compared to the reference. The reason why the KB-IBI method (solid red line in Fig. 5) does a better job in the reproduction of the potential is because the correction term shifts the potential directly by adding a ramp potential on top of it (Eq. (9)). In C-IBI the correction has not direct relation to the potential, since the update is based on matching the integrated RDF over the whole range.

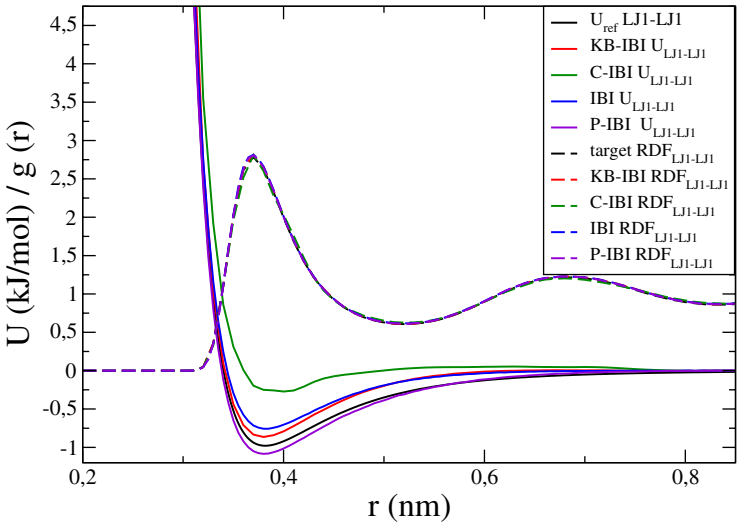


Fig. 5. Comparison of the different IBI methods with the atomistic reference (black) for the LJ1-LJ1 interaction, RDFs (dashed lines) and final potentials (solid lines) for KB-IBI(red), C-IBI (green), IBI (blue) and P-IBI (violet).

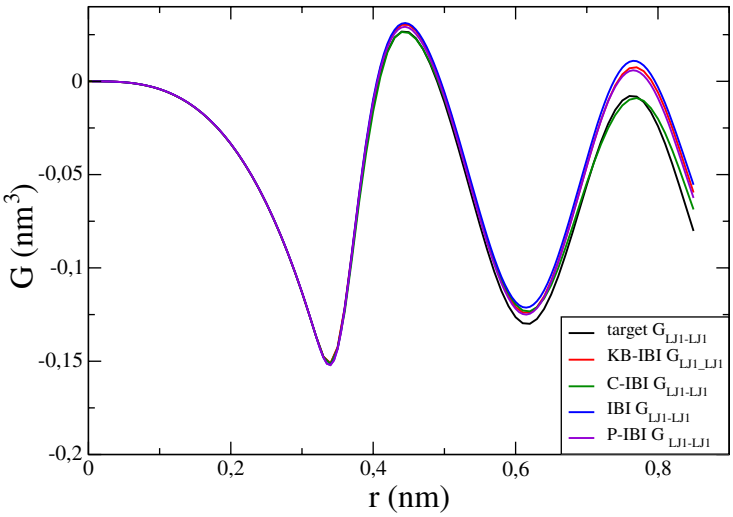


Fig. 6. Comparison of the RKBIs for the LJ1-LJ1 interaction: reference (black), KB-IBI (red), C-IBI (green), IBI (blue) and P-IBI (violet).

To summarize it, the C-IBI scheme reproduces the target RKBI best by construction. This is achieved, however, with a potential which deviates significantly from the reference atomistic potential. Potentials obtained from KB-IBI, P-IBI and IBI do not match the target RKBI in the long distance range but approximate the reference atomistic potential better. Agreement of, both, the pair potential and the RKBI with the atomistic references cannot be achieved with any of the IBI-related methods for the binary Lennard-Jones system studied in this work. Neither the KB-IBI method nor the C-IBI method provide an improvement over the P-IBI method in terms of the reproduction of the reference potential. We note that for the present Lennard-Jones system a complication arises from the fact that the RKBIs do not converge to a

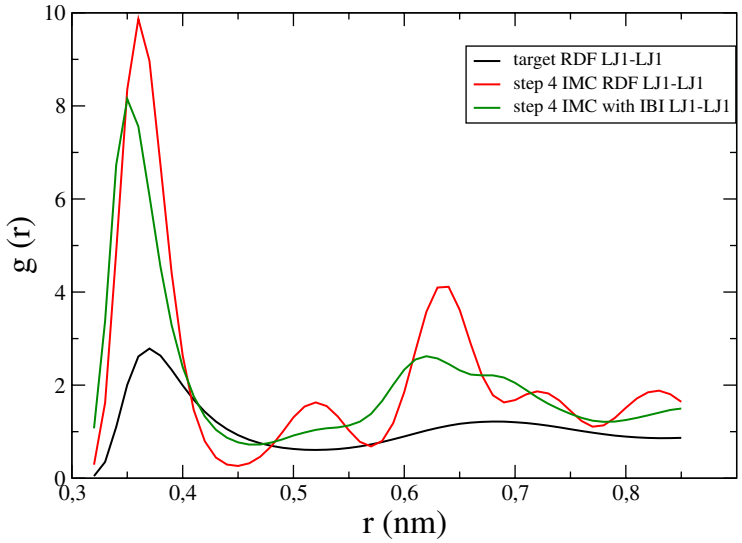


Fig. 7. Target RDF between LJ1-LJ1 (black line), RDF after 4 IMC steps without previous IBI (red line), RDF after 4 steps with previous IBI (green line).

limiting plateau value but keep on oscillating at long range. This makes it difficult to apply the KB-IBI update (see Eq. (9)), which relies on a limiting plateau value being found in the RKBI.

4.2 Effective pair potentials via the inverse Monte-Carlo method

Since the goal of IMC is the same as for IBI, i.e. generation of effective pair potentials by achieving a match in the RDF between the CG and FG system, comparable results are expected.

As we see from the red curve in Fig. 7, simple application of IMC leads to a complete mismatch in the RDF and a non-physical shape of the potential (Fig. 8). This results in a crash of the simulation after a few iterations. It is known that this is likely to happen if the initial guess, i.e. the PMF, is too far away from the solution of the problem. To provide a better starting point 30 preliminary steps of P-IBI are performed. But from the green curve in Figs 7 and 8 we do not see any further improvement. The algorithm cannot converge to a stable solution. For further stabilization we add a regularization term to the IMC update scheme according to Eq. (14).

A singular value decomposition (SVD) of the matrix A (see Eq. (13)) allows an analysis on how large the regularization parameter λ has to be. The SVD decomposes the matrix A to

$$A = U\Sigma V^T = \sum_{k=1}^{\mathbf{n}} \mathbf{u_k}\sigma_k \mathbf{v_k^T} \tag{18}$$

where Σ is a diagonal matrix with the singular values $\sigma_1, ..., \sigma_n$ of A and $V(\mathbf{v_1}, ...\mathbf{v_n})$ and $U(\mathbf{u_1}, ...\mathbf{u_n})$ are both unitary $n \times n$ matrices. Now substitution of Eq. (18) in Eq. (15) leads to

$$\Delta U_{ij} = \sum_{k=1}^{n} \frac{\sigma_k}{\sigma_k^2 + \lambda} \left(\boldsymbol{u_k^T}(g_{ij}^{n-1} - g_{ij}^{ref})\right)\mathbf{v_k}. \tag{19}$$

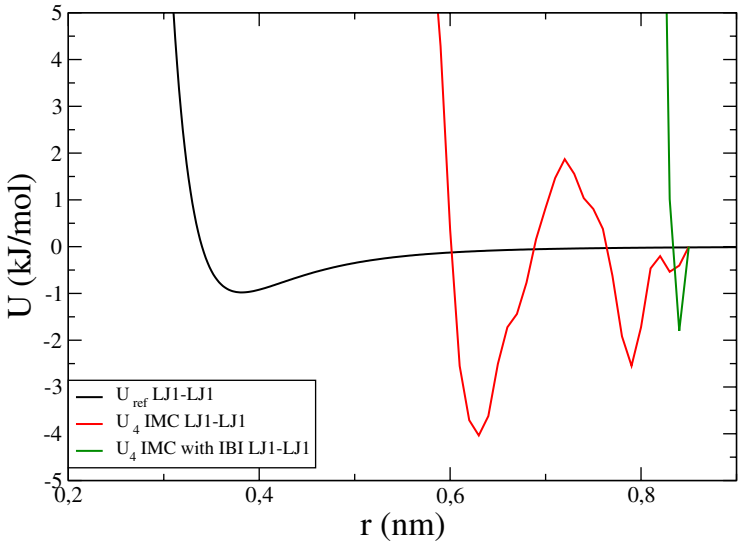


Fig. 8. Reference LJ potential for the LJ1-LJ1 interaction (black line), $U_{LJ1-LJ1}$ after 4 IMC steps without previous IBI (red line), $U_{LJ1-LJ1}$ after 4 steps with previous IBI (green line).

A good regularization parameter λ should be such that it dominates the smallest singular values (squared) but is itself small compared to the larger ones. The optimal λ parameter can only be determined if the exact solution is known, which unfortunately is not the case in practice. Finding good heuristic choices of λ is still a problem that is intensively studied in the mathematical community.

In our case the SVD analysis shows the smallest singular values of a magnitude of 100 or below. These are the values we considered to be regularized. So the IMC method is run with a regularization parameter λ of 100. The red curves in Fig. 9 show the final result of 4 IMC updates per interaction with a regularization value of 100. As one can see with additional regularization a match in the RDF could be achieved and a physical correct looking shape of the potential was obtained for all three interactions LJ1-LJ1 (upper panel), LJ2-LJ2 (middle panel) and LJ1-LJ2 (lower panel). Nevertheless the generated potentials are a bit too attractive compared to the reference potentials (black lines in all the panels of Fig. 9). That is why larger regularization parameter of 300 and 1000 are tested. These larger values lead to a better approximation towards the reference potential (green lines $\lambda = 300$ and blue lines $\lambda = 1000$ in Fig. 9). Especially with $\lambda = 1000$ the potentials for the LJ1-LJ1 and LJ1-LJ2 interaction could almost be fully recovered. Both potentials show only minor divergence in the long-range part of the potential. Beyond that the generated LJ1-LJ2 potential has a slightly deeper potential well. The generated LJ2-LJ2 potential stays a bit more attractive for both larger λ values.

For further calculations we take the potentials generated with a regularization parameter of 1000. Although this value might lead to a too strong dominance of the regularization term over the data fit term in Eq. (14), it leads to the best reproduction of the underlying potentials for all three interactions.

In this section we could show that the run of previous IBI steps and addition of a regularization term stabilizes the IMC method. Without these auxiliary steps the algorithm did not converge to a stable solution. But with them, comparable results to the P-IBI method are obtained as shown in Fig. 10. In this graph the results of P-IBI after 80 steps and the regularized IMC run ($\lambda = 1000$) after 4 steps (with preliminary

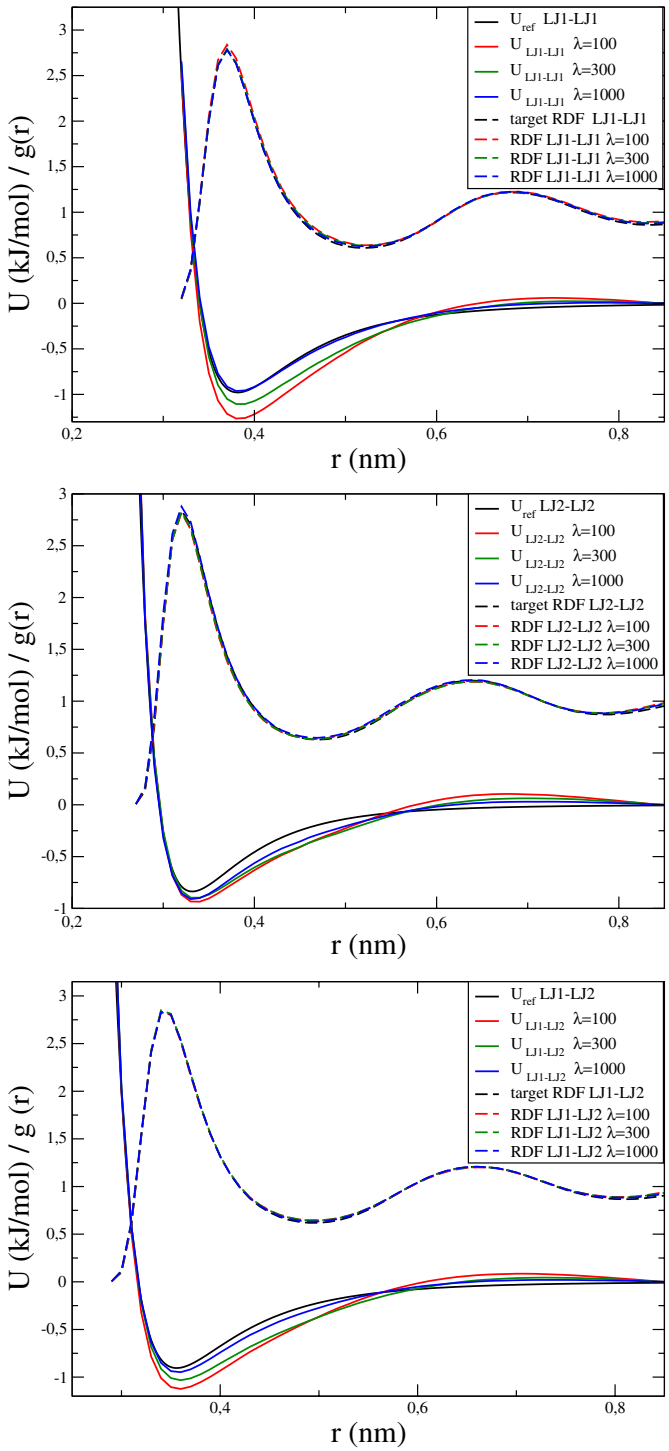


Fig. 9. Influence of different regularization values on the IMC results shown for the LJ1–LJ1 (upper panel), LJ2–LJ2 (middle panel) and LJ1–LJ2 (lower panel) interaction: Potentials (solid lines) and RDFs (dashed lines) for the FG system (black), for $\lambda = 100$ (red), $\lambda = 300$ (green) and $\lambda = 1000$ (blue).

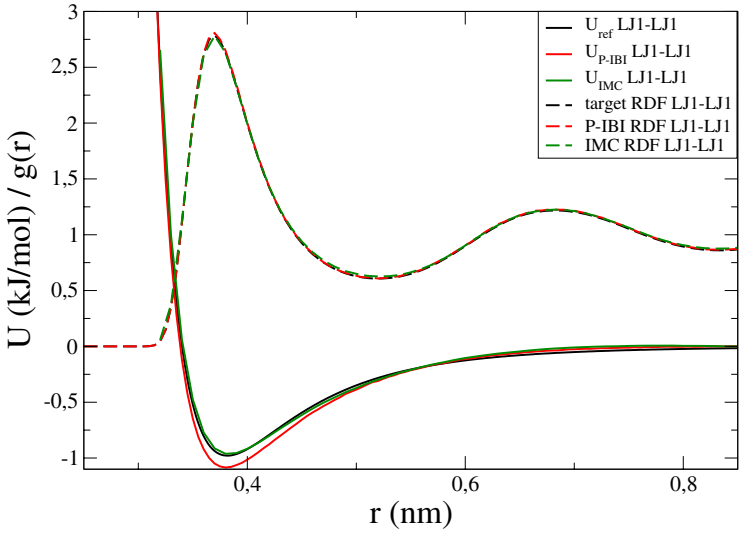


Fig. 10. Final potentials (solid lines) and RDFs (dashed lines) of the LJ1–LJ1 interaction obtained from P-IBI (red) and IMC with a regularization parameter of 1000 (green) in comparison with the FG reference (black).

30 P-IBI steps) for the LJ1-LJ1 interaction are compared. One sees clearly that the IMC method results in a better reproduction of the reference Lennard-Jones potential than the P-IBI method does. This also holds true for the LJ1-LJ2 interaction. The final LJ2-LJ2 interaction is for both methods nearly the same, but in both cases slightly too attractive. Overall it seems that the interdependence of the potential update in IMC leads to an improvement in the reproduction of all potentials present in a binary system. Compared to P-IBI also less iterations are needed to obtain reasonable well matched RDFs and potentials. But as there is no free lunch, the fewer amount of iterations comes with the cost of a longer run time per iteration.

4.3 Thermodynamic quantities

The generated effective pair potentials are now further evaluated as described in Sect. 3.3. Before we start with the KB analysis, we first investigate how well the RDFs are reproduced for the different potentials. In Fig. 11 we see the RDFs for the LJ1–LJ1 interaction in good agreement with the FG reference (target RDF) even beyond the cut-off for mostly all of the applied methods. Only the KB-IBI derived potentials lead to a small shift in the RDF towards smaller distances. The LJ2–LJ2 RDFs (not shown here) obtained with IMC and KB-IBI potentials show a slightly larger first peak, whereas slightly smaller first and second maxima are obtained with the C-IBI derived potential. For the LJ1–LJ2 interaction the RDF (not shown here) obtained with the KB-IBI potential shows a first and second maximum which are too large, whereas the other methods are in good agreement with the reference. So in general one can say that an update based on the KB-IBI method comes with the cost of less structural accuracy in the production run. This structural inaccuracy might be related to the fact, that KB-IBI targets to match a quantity which is mainly determined by the long-range part of the potential, whereas the RDF is mainly determined by the short-range part of the potential. Thus to get accurate structural representation, it may be preferable not to match the reference potential exactly. It may be more

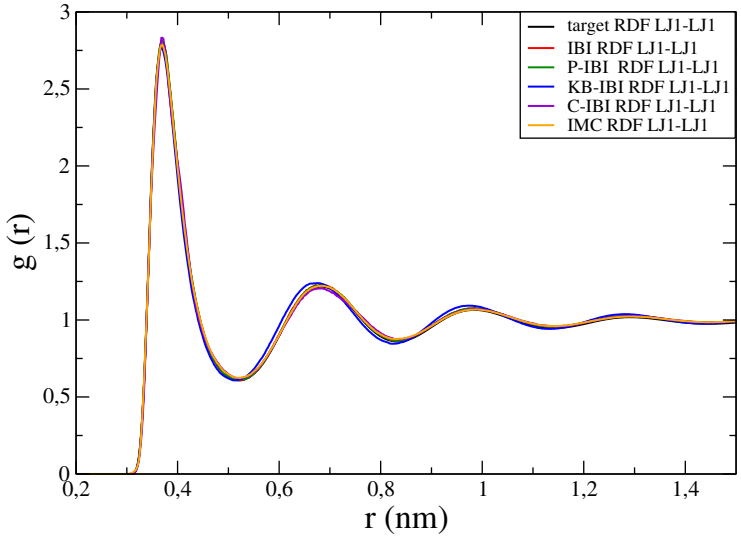


Fig. 11. RDF for the LJ1-LJ1 interaction obtained from different generating potentials: reference LJ potential (black), IBI (red), P-IBI (green), KB-IBI (blue), C-IBI (violet) and IMC (orange).

important that the update of the generating method focuses on the short-range part of the potential, since this is the part which mainly is determining the RDF.

In order to calculate thermodynamic quantities using KBIs obtained from molecular simulations, finite size artefacts should be considered. To this end, Ganguly et al. applied a tail correction to the RDF and studied the dependence of the RKBI on system size in NpT simulations [48]. They could show that for sufficiently large systems, the RKBI approaches a limiting plateau value, which corresponds to the correct thermodynamic limit (Eq. (7)). In particular, this could be demonstrated by comparing the KBIs obtained by means of this approach with an alternative method introduced by Schnell et al. [62,63]. These authors showed that KBI values in the thermodynamic limit can be obtained from an analysis of particle number fluctuations in small sub-volumes (sub-boxes) $V_s \subset V$ in simulations of closed NVT (or NpT) systems. By following the formalism of Hill, [64] they could show that SKBIs ($G_{ij}^{V_s}$, defined below) scale linearly with the inverse of system size $L_s = V_s^{1/3}$ of these small subboxes [62]:

$$G_{ij}^{V_s} = G_{ij}^{\infty} + \frac{a}{L_s} \tag{20}$$

The small boxes can be considered as grand-canonical ensembles embedded in a larger box, which acts as particle bath with which an exchange of particles and energy takes place. Assuming our simulation box (NVT or NpT) with size L is large enough compared to the small box with size L_s, this method can be applied for our system as well. The SKBIs (see Eq. (20)) are calculated from particle number fluctuations in these small sub-boxes according to [63]:

$$G_{ij}^{V_s} = V_s \left[\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_i \rangle} \right] \tag{21}$$

where N_i is the particle number of type i in the sub-box with the volume V_s and δ_{ij} is the Kronecker delta. The averages are taken over sub-boxes randomly inserted along the trajectory of the NVT system. Following this procedure, SKBIs are calculated for

Table 3. Limiting KBI values for the interactions LJ1-LJ1, LJ2-LJ2, LJ1-LJ2 for different generating potentials.

Method	G_{11}^{∞} (nm^3)	G_{22}^{∞} (nm^3)	G_{12}^{∞} (nm^3)
atomistic	−0.054	−0.029	−0.032
IBI	−0.055	−0.031	−0.031
P-IBI	−0.055	−0.030	−0.030
KB-IBI	−0.054	−0.023	−0.034
C-IBI	−0.063	−0.045	−0.021
IMC	−0.054	−0.025	−0.033

various sub-box sizes (L_s). From these calculations the KBIs in the thermodynamic limit can be obtained via Eq. (20). To make use of this equation, first the regime where the SKBIs (Eq. (21)) scale linear with the inverse box size has to be identified and second this regime has to be extrapolated for $L_s \to \infty$ or $1/L_s \to 0$. In table 3 the limiting KBI values (G_{ij}^{∞}) for all interactions with different generating potentials are presented. In Fig. 12 the plots from which the limiting KBI values are obtained are shown. Since we choose a spherical shape for the sub-boxes, the radii of these spheres is used as size L_s.

In table 3 the limiting KBIs for the different generating potentials are listed. The KBIs obtained from potentials derived via IBI and P-IBI slightly underestimate the atomistic reference for the LJ1-LJ1 and LJ2-LJ2 interaction. They overestimate the LJ1-LJ2 KBI. The potentials derived via KB-IBI and IMC show the best reproduction of the KBI for the LJ1-LJ1 interaction. They underestimate the KBI of the LJ2-LJ2 interaction and overestimate the KBI for the LJ1-LJ2 interaction. It is interesting to see that the C-IBI method shows the poorest performance in reproducing the KBIs for all interactions, although this method shows the best fit in case of the RKBIs. A reason for this mismatch might be that finite size effects have a too strong influence on the potentials derived via C-IBI. This also applies for the KB-IBI method. But in case of KB-IBI the underlying potential is closer to the reference one than it is the case for C-IBI. This might lead to the better representability of the KBI by potentials derived via KB-IBI compared to potentials generated via C-IBI.

Having obtained the limiting KBI values, we calculate the isothermal compressibility, κ_T, of the system and the partial molar volume for each of the particles, $\overline{v_1}$ and $\overline{v_2}$. Following the KB theory the compressibility can be calculated as [36]

$$\kappa_\tau = \frac{\zeta}{\eta\ k_B T} \tag{22}$$

and the partial molar volume for particle type LJ1 as [36]

$$\overline{v_1} = \frac{1 + \rho_2(G_{22}^{\infty} - G_{12}^{\infty})}{\eta} \tag{23}$$

and for type LJ2 respectively as [36]

$$\overline{v_2} = \frac{1 + \rho_1(G_{11}^{\infty} - G_{12}^{\infty})}{\eta} \tag{24}$$

where

$$\zeta = 1 + \rho_1 G_{11}^{\infty} + \rho_2 G_{22}^{\infty} + \rho_1\rho_2(G_{11}^{\infty}G_{22}^{\infty} - G_{12}^{\infty 2}), \tag{25}$$

and

$$\eta = \rho_1 + \rho_2 + \rho_1\rho_2(G_{11}^{\infty} + G_{22}^{\infty} - 2G_{12}^{\infty}), \tag{26}$$

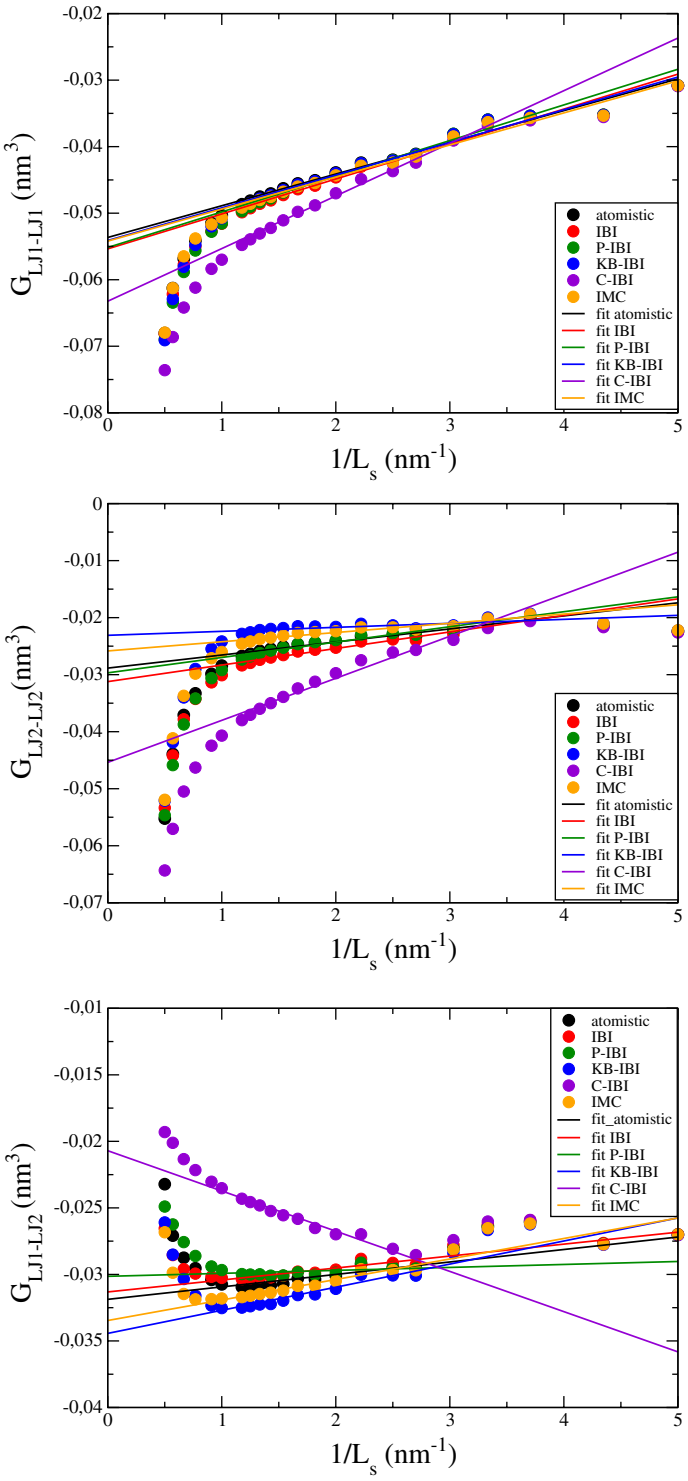


Fig. 12. LJ1-LJ1 (upper panel), LJ2-lJ2 (middle panel) and LJ1-LJ2 (lower panel) SKBIs calculated from number fluctuations of spherical sub-boxes as a function of the inverse sub-box radii for the differently generated potentials with different system size (solid dots). Solid straight lines are fitted to the linear regimes of the plots.

Table 4. Virial pressure p in [bar], the isothermal compressibility κ_τ in $[10^{-9}\,\mathrm{m}^2/\mathrm{N}]$, the partial molar volumes $\overline{v_1}$ and $\overline{v_2}$ in $[\mathrm{nm}^3]$, the derivative of the activity coefficient $(\frac{\partial ln\gamma_i}{\partial ln x_i})_{p,T}$ and the total energy density in $[10^{-22}\,\mathrm{kJ/nm}^3]$ for the different generating potentials

Method	p	κ_τ	$\overline{v_1}$	$\overline{v_2}$	$\left(\frac{\partial ln\gamma_1}{\partial ln x_1}\right)_{p,T}$	Energy density
atomistic	1.0	1.07	0.0298	0.0208	0.133	−1.762
IBI	942.4	0.83	0.0296	0.0207	0.176	−0.836
P-IBI	16.7	1.31	0.0298	0.0204	0.182	−1.739
KB-IBI	609.8	1.02	0.0315	0.0208	0.055	−1.135
C-IBI	2317.9	0.71	0.0256	0.0172	0.726	0.321
IMC	413.1	0.99	0.0308	0.0208	0.089	−1.442

ρ_1 is the number density for particle type LJ1, G_{11}^{∞} is the limiting KBI for the LJ1-LJ1 interaction, ρ_2 is the number density for particles type LJ2, G_{22}^{∞} is the limiting KBI for the LJ2-LJ2 interaction and G_{12}^{∞} is the limiting KBI for the LJ1-LJ2 interaction [36].

To complete the KB analysis, the derivative of the activity coefficient, γ_1, with respect to the mole fraction of the mixture, x_1, for particle type LJ1 is calculated [65].

$$\left(\frac{\partial \ln \gamma_1}{\partial \ln x_1}\right)_{p,T} = -\frac{\rho_2 x_1 (G_{11}^{\infty} + G_{22}^{\infty} - 2G_{12}^{\infty})}{1 + \rho_2 x_1 (G_{11}^{\infty} + G_{22}^{\infty} - 2G_{12}^{\infty})}. \tag{27}$$

The results for the KB analysis together with the virial pressure and the average energy density of the system are presented in table 4.

As one sees from table 4 the reference pressure of 1 bar is not matched with any of the derived potentials. The P-IBI potentials provide the best approximation of the reference value, but in this case the compressibility is slightly overestimated compared to the atomistic reference. C-IBI potentials generate the highest virial pressure compared to all other potentials which is due to a mismatch of the C-IBI potentials in the attractive and repulsive regions (see Fig. 5). In case of the IBI method, the potential well is too high which also leads to more repulsion between the particles and as a consequence to a larger virial pressure. The potentials generated via IMC also overestimate the virial pressure despite that all three IMC potentials are closest to the atomistic reference potentials. This may be explained because IMC potentials are slightly more repulsive in the tail compared to reference potential (see Fig. 9). This shows that even small differences in the potential have a big impact on the pressure of the system. Moreover it shows how well the long range part of the potential has to be matched to achieve the correct pressure. The KB-IBI potentials also overestimate the virial pressure of the system, but reproduce the compressibility of the system quite well compared to the atomistic reference. In general, the compressibility is underestimated if the potential is too repulsive compared to the atomistic reference. The partial molar volumes of the two different components agree well with the reference values for nearly all potentials, except the C-IBI generated potentials. This results from the poor agreement of the limiting KBIs with the atomistic reference as we can see from Fig. 12. The derivatives of the activity coefficient overestimate the reference values for mostly all IBI potentials. The IMC and KB-IBI potentials however underestimate this quantity.

All the results of the KB analysis can be related to how well each method could reproduce the limiting KBIs compared to the target system. A weakness of the small-system method applied to calculate the limiting KBIs is the reliability of the linear fit, especially if the differences have a small magnitude. But from the thermodynamic analysis we see that properties that are insensitive to the long range part of the potential can be reproduced by all applied methods in good agreement with the

atomistic reference. To achieve a better match in a quantity, e.g. the derivative of the activity coefficient, that is mainly determined by the tails of the potential, a better reconstruction of this part for all interactions in the system has to be guaranteed.

Finally the average energy densities of the systems have been compared. This quantity reflects how well the different methods could reproduce the potentials for all interactions present in the system. P-IBI potentials reproduce the energy density most accurately. Here all the generated potentials have a slightly deeper minimum compared to the atomistic reference potentials. KB-IBI and IMC potentials provide less negative energy densities. A reason for this can be the more repulsive tails of the potentials derived via IMC. KB-IBI potentials are also always slightly more repulsive compared to the atomistic reference, but less repulsive than potentials generated via IBI. That is why in this case the energy density is even less negative. The energy density predicted by the C-IBI potentials is positive, which can again be related to the overly repulsive potentials generated by this method.

5 Conclusion

In this paper we have addressed the question whether it is possible to recover the generating potentials of target RDFs via different variations of the IBI method and the IMC method. The point of interest was to see if these methods can converge to a known solution of the inverse problem of finding effective pair potentials for non-bonding interactions derived from a given set of RDFs. The system under study was a binary mixture of two LJ particles.

For none of the methods it was possible to recover the potentials for all interactions, although good structural match could be achieved for all methods that have been considered. We could show that for IBI-type methods implementation of thermodynamic constraints improves the convergence of the method with respect to the potential and the RDF, if used in combination with a well chosen cut-off. The cut-off should be chosen such that it is large enough to include the relevant parts of the potential, but not too large, since this affects the convergence negatively. As additional thermodynamic constraints the pressure (P-IBI), KBIs (KB-IBI) and the integrated RDF (C-IBI) have been investigated. An update based on C-IBI does not improve the convergence towards the reference potential most probably because of finite size effects. The KB-IBI method leads to results that are in between those obtained with P-IBI and C-IBI. The linear corrections used in P-IBI and KB-IBI lead to reproducing the reference potential better than the alternative update based on the integrated RDF used in C-IBI. The C-IBI-based update however provides the best representation of the RKBIs. In terms of convergence of the method towards the reference potential, inclusion of the pressure as a thermodynamic constraint in the standard IBI procedure is preferred over the alternative quantities (KBIs or integrated RDFs), although this may interfere with the Henderson theorem. [17]

For IMC we could show how sensitive this method reacts to several factors. First, the initial guess should be close to the actual potential. Hence, the PMF is not a good choice. Second, one should provide enough statistics to achieve convergence of the algorithm. In our case we could not achieve convergence within $5 \cdot 10^6$ MD steps (compare IBI: $1 \cdot 10^6$ MD steps). But we could show that it is possible to decrease the statistical demand by adding a regularization term, which stabilizes the algorithm. A SVD analysis of the matrix helps to get a good educated guess for the magnitude of the regularization parameter. The regularization allows the generation of potentials which have the best agreement with the reference potentials for all three interactions present in the binary system.

We further addressed the thermodynamic representability of the derived potentials in comparison with the atomistic reference. To this end, the isothermal compressibility, the partial molar volumes of the binary mixture components, and the derivative of the activity coefficient with respect to the mole fraction (for component 1) have been calculated from the KBIs obtained with the derived potentials. The IBI potentials do not reproduce the isothermal compressibility of the atomistic reference model. The KB-IBI potentials best reproduce this quantity. The partial molar volumes of the two mixture components are represented well with the majority of the potentials. The pressure of the system is overestimated with all potentials. While the P-IBI potential shows the best agreement with the reference pressure, it overestimates the isothermal compressibility. All IBI potentials overestimate the derivative of the activity coefficient computed for one of the components. Hence, these potentials do not accurately represent the dependence of the chemical potential on composition which is mainly determined by the tails of the LJ potentials. The KB-IBI potentials underestimate the derivative of the activity coefficient. The reason that KB-IBI and C-IBI do not well represent the quantities they are designed for is related to finite-size system effects, which are stronger in Lennard-Jones mixtures as compared to aqueous solution mixtures previously studied [47,49].

Comparison with the results obtained with IMC potentials further showed that even a better match of the potential around the attractive minimum does not improve the thermodynamic representability of the model. The non-sensitivity of the RDFs on long-range tails of the potentials clearly poses a challenge in representing thermodynamic properties with models derived by inverse coarse-graining methods. The derived potentials that best reproduce the attractive tail of the reference potential also represent the average energy density of the system better, but already small deviations in the long range part do not allow an accurate representation of this quantity.

The authors would like to thank Christine Peter, Christoph Junghans, Denis Andrienko, Christoph Scherer and Pritam Ganguly for fruitful discussions and useful remarks on technical aspects of the applied methods. Funding has been granted by the German Research Foundation (DFG) within the Collaborative Research Center "Multiscale Simulation Methods for Soft-Matter Systems" (SFB-TRR 146). Computational time provided on the Lichtenberg High Performance Computer by the TU Darmstadt is acknowledged.

References

1. W. Tschöpp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn, Acta Polym. **49**, 61 (1998)
2. M. Murat, K. Kremer, J. Chem. Phys. **108**, 4340 (1998)
3. F. Müller-Plathe, ChemPhysChem **3**, 754 (2002)
4. M. Praprotnik, L. Delle Site, K. Kremer, Annu. Rev. Phys. Chem. **59**, 545 (2008)
5. C. Peter, K. Kremer, Soft Matter **5**, 4357 (2009)
6. E. Brini, E.A. Algaer, P. Ganguly, C. Li, F. Rodriguez-Ropero, N.F.A. van der Vegt, Soft Matter **9**, 2108 (2013)
7. T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, M. Karttunen, Phys. Chem. Chem. Phys. **11**, 1869 (2009)
8. B. Hess, C. Holm, N.F.A. van der Vegt, J. Chem. Phys. **124**, 164509 (2006)
9. Y. Wang, W.G. Noid, P. Liu, G.A. Voth, Phys. Chem. Chem. Phys. **11**, 2002 (2009)
10. E. Brini, V. Marcon, N.F.A. van der Vegt, Phys. Chem. Chem. Phys. **13**, 10468 (2011)
11. D. Reith, M. Puetz, F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003)
12. A.P. Lyubartsev, A. Laaksonen, Phy. Rev. E **52** (4), 3730 (1995)
13. F. Ercolessi, J.B. Adams, Europhys. Lett. **26**, 583 (1994)
14. S. Izvekov, G.A. Voth, J. Phys. Chem. B **109**, 2469 (2005)

15. J.W. Mullinax, W.G. Noid, J. Phys. Chem. C **114**, 5661 (2010)
16. M.S. Shell, J. Chem. Phys. **129**, 144108 (2008)
17. R.L. Henderson, Phys. Lett. **49A**, 197 (1974)
18. W G. Noid, J.-W. Chu, G.S. Ayton, V. Krishna, S. Izvekov, G.A. Voth, A. Das, H.C. Andersen, J. Chem. Phys. **128**, 244114 (2008)
19. W.G. Noid, P. Liu, Y. Wang, J.-W. Chu, G.S. Ayton, S. Izvekov, H.C. Andersen, G.A. Voth, J. Chem. Phys. **128**, 244115 (2008)
20. M.E. Johnson, T. Head-Gordon, A.A. Louis, J. Chem. Phys. **126**, 144509 (2007)
21. E. Brini, N.F.A. van der Vegt, J. Chem. Phys. **137**, 154113 (2012)
22. P. Ganguly, N.F.A. van der Vegt, J. Chem. Theory. Comput. **9**, 5247 (2013)
23. L.C. Jacobson, R.M. Kirby, V. Molineo, J. Phys. Chem. B **118**, 8190 (2014)
24. H. Wang, C. Junghans, K. Kremer, Eur. Phys. J. E. **28**, 221 (2009)
25. A. Das, H.C. Andersen, J. Chem. Phys. **132**, 164106 (2010)
26. N.J. Dunn, W.G. Noid, J. Chem. Phys. **143**, 243148 (2015)
27. B. Hess, C. Holm, N.F.A. van der Vegt, J. Chem. Phys. **124**, 164509 (2006)
28. J.W. Shen, C. Li, N.F.A. van der Vegt, C. Peter, J. Chem. Theory. Comput. **7**, 1916 (2011)
29. H.J. Qian, P. Carbone, C. Xiaoyu, H.A. Karimi-Varzaneh, C.C. Liew, F. Müller-Plathe, Macromolecules **41**, 9919 (2008)
30. H. Eslami, H.A. Karimi-Varzaneh, F. Müller-Plathe, Macromolecules **44**, 3117 (2011)
31. M. Langeloth, T. Sugii, M.C. Boehm, F. Müller-Plathe, J. Chem. Phys. **143**, 243158 (2015)
32. C. Peter, L. Delle Site, K. Kremer, Soft Matter **4**, 859 (2004)
33. S. Jain, S. Garde, S.K. Kumar, Ind. Eng. Chem. Res. **45**, 5614 (2006)
34. C. Fu, P.M. Kulkarni, M.S. Shell, L.G. Leal, J. Chem. Phys. **137**, 164106 (2012)
35. D. Ivanizki, *Numerical Analysis of the relation between interactions and structures in a molecular fluid* (Ph.D Thesis University of Mainz, 2015).
36. A. Ben-Naim, *Molecular Theory of Solutions* (Oxford University Press New York, 2006).
37. J.G. Kirkwood, F.P. Buff, J. Chem. Phys. **19**, 774 (1951)
38. P. Krüger, S.K. Schnell, D. Bedeaux, S. Kjelstrup, T.J.H. Vlugt, J.-M. Simon, J. Phys. Chem. Lett. **4**, 235 (2013)
39. S. Weerashinge, P.E. Smith, J. Chem. Phys. **118**, 10663 (2003)
40. S. Weerashinge, P.E. Smith, J. Phys. Chem. B **107**, 3891 (2003)
41. S. Weerashinge, P.E. Smith, J. Chem. Phys. **121**, 2180 (2004)
42. S. Weerashinge, P.E. Smith, J. Phys. Chem. B **109**, 15080 (2005)
43. M.E. Lee, N.F.A. van der Vegt, J. Chem. Phys. **122**, 114509 (2005)
44. M.B. Gee, N.R. Cox, Y.F. Jiao, N. Bentenitis S. Weerashinge, P.E. Smith, J. Chem. Theory Comput. **8**, 1802 (2012)
45. M. Kang, P.E. Smith, J. Comput. Chem. **27**, 1477 (2006)
46. M. Fyta, R.R. Netz, J. Chem. Phys. **136**, 124103 (2012)
47. P. Ganguly, D. Mukherji, C. Junghans, N.F.A van der Vegt, J.Chem. Theor. and Comp. **8**, 1802 (2012)
48. P. Ganguly, N.F.A. van der Vegt, J. Chem. Theory. Comput. **9**, 1347 (2013)
49. T.E. Oliveira, P.A. Netz, K. Kremer, C. Junghans, D. Mukherji, J. Chem. Phys. **144**, 174106 (2016)
50. A. Lyubartsev, A. Mirzoev, L-J. Chen, A. Laaksonen Faraday Discuss. **144** 43 (2010)
51. M. Hanke, arXiv:1603.03899 [math-ph] (2016)
52. M. Hanke, arXiv:1603.03900 [math-ph] (2016)
53. T. Murtola, E. Falck, M. Karttunen, I. Vattulainen, J. Chem. Phys. **126**, 075101 (2007)
54. H.W. Engl, M. Hanke, A. Neubauer *Regularization of Inverse Problems* (Kluver Academic Publishers, 2000).
55. V. Ruehle, C. Junghans, A. Lukyanov, K. Kremer, D. Andrienko, J. Chem. Theo. Comp. **5**, 3211 (2009)
56. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theo. and Comp. **4**, 435 (2008)

57. V. Ruehle, C. Junghans, Macromol. Theory Simul. **20**, 472 (2011)
58. S.Y. Mashayak, M. Jochum, K. Koschke, N.R. Aluru, V. Ruehle, C. Junghans, PLoS one **10**, e131754 (2015)
59. M. Parrinello, J. Appl. Phys. **52**, 7182 (1981)
60. S. Nosė. M.L. Klein, Mol. Phys. **50**, 1055 (1983)
61. `https://github.com/votca/downloads/blobmaster/votca-csg-manual-1.3_rc1.pdf`
62. S.K. Schnell, T.J.H. Vlugt, J.-M. Simon, D. Bedeaux, S. Kjelstrup, Chem. Phys. Lett. **504**, 199 (2011)
63. S.K. Schnell, X. Liu, J.-M. Simon, A. Bardow, D. Bedeaux, T.J.H. Vlugt, S. Kjelstrup, J. Phys. Chem. B **115**, 10911 (2011)
64. T.L. Hill, *Thermodynamics of Small Systems, Part 1* (W. A. Benjamin: New York, 1963).
65. A. Villa, C. Peter, N.F.A. van der Vegt, J. Chem. Theory. Comput. **6**, 2434 (2010)

The relative entropy indicates an ideal concentration for structure based coarse graining of binary mixtures

David Rosenberger* and Nico F. A. van der Vegt
Eduard Zintl Institut für Anorganische und Physikalische Chemie,
Technische Universität Darmstadt, Darmstadt, Germany

Many methodological approaches have been proposed to improve systematic or bottom-up coarse graining techniques to enhance the representability and transferability of the derived interaction potentials. Transferability describes the ability of a CG model to be predictive, i.e. to describe a system at state points different from those chosen for parametrization. Whereas the representability characterizes the accuracy of a CG model to reproduce target properties of the underlying reference or fine grained (FG) model at a given state point. In this article, we shift the focus away from methodological aspects and rather raise the question if we can overcome the disadvantages of a given method in terms of representability and transferability by systematically selecting the state point at which the CG model gets parametrized. We answer this question by applying the Inverse Monte Carlo (IMC) approach - a structure based coarse graining method - to derive effective interactions for binary mixtures of simple Lennard-Jones (LJ) particles, which are different in size. For such simple systems we indeed can identify a concentration where the derived potentials show the best performance in terms of structural representability and transferability. This specific concentration is identified by computing the relative entropy which quantifies the information loss between different IMC models and the reference LJ model at varying mixture compositions. Further, we show that an IMC model for mixtures of n-hexane and n-perfluorohexane shows the same trend in transferability as the IMC models for the LJ system. All derived models are more transferable in the direction of increasing concentration of the larger sized compound.

Pages: 51-83:

* rosenberger@cpc.tu-darmstadt.de

I. INTRODUCTION

Limitations of all atom based molecular dynamics (MD) simulations to time and length scales of a microseconds and a few nanometers impede an accurate modeling of soft matter systems. Therefore, the development of simplified models has been fostered over the last decades. These models consist of a fewer number of particles and have less degrees of freedom available than their all atom reference.[1–4] The projection of a high resolution or fine grained (FG) model onto a configuration space lower in resolution is commonly referred to as coarse graining.[5–7] Coarse grained (CG) models can be generated either in a bottom-up or in a top-down approach.[8, 9] While the latter one relies on parameterizing an analytical function to describe the interactions present in the system[10, 11], bottom-up CG models are derived based on information obtained from an underlying FG model. Thus, the advantage of bottom-up coarse graining is the possibility to maintain specific features of the FG model, which cannot be captured by fitting thermo-physical properties directly in top-down coarse graining. Multiple methods exists to generate CG models in a bottom up manner. The conditional reversible work method[12] or the effective force coarse graining approach[13], among others[14–17], are methods to derive a CG model by computing pair forces between particles without taking higher multibody correlations into account. Contrary to those derived coarse graining method are so called inverse methods. Here, higher order correlations play a crucial role as multibody effects are mapped onto effective pair forces such that the CG model has a minimal difference to the underlying FG system with respect to a specific target property. This target property can be the total force on each particle[18], as done in multiscale coarse graining (MS-CG)[19–21], or the pair structure, as in iterative Boltzmann inversion[22, 23] or inverse Monte Carlo (IMC)[24] or higher order structural correlation functions.[25–28]

Despite this great variety of methods, the key challenge for all approaches is to achieve an accurate description of the FG model at a CG level (representability) and the applicability of the CG model at state points different than the one chosen during parametrization (transferability). It is difficult to achieve both simultaneously as a theoretical relation between representability and transferability is missing.[29–31] What is known is that embedding of higher order correlations in the effective pair potentials lead to a very accurate representability of the specific target quantity chosen for parametrization, but not of others as well.[32–35] More importantly, higher order correlations inherent in CG models make optimized pair potentials barely transferable, since they are characteristic for a given state point.[36] On the other hand, excluding those correlations in the parametrization procedure leads to more transferable, but less representable CG models, especially in terms of structure.[13, 37–39] This illustrates that the choice of a specific coarse graining method always leads to a com-

promise which one has to make between the representability and transferability of the final potentials.

In recent years, several approaches have been proposed to find better compromises. The addition of linear ramp potentials or the application of a constrained version of a method improved the thermodynamic representability of the derived CG models as well as their transferability.[23, 40–43] Further, parameterization via an extended ensemble approach or the use of an extended Hamiltionian description for the CG system could improve the model's transferability.[44–52]

Another critical parameter for coarse graining is the selection of the mapping scheme. This scheme defines how the CG particles are constructed and it can have a significant influence on the performance of a CG model. [27, 53] But despite recent efforts, a systematic way to select an optimal mapping scheme is still not feasible.[54–56]

Being aware of the limitations which may arise either from the coarse graining method or from the mapping or from both, in the present work we want to focus on an aspect which has been barely studied. This aspect is the systematic choice of a state point to parametrize CG models. Inherent to this aspect is the question: can we overcome deficits arising from a chosen method by systematically selecting a reference state point? To answer this question we apply a structure based coarse graining method, namely IMC, to construct effective pair potentials for binary mixtures of simple Lennard-Jones particles different in size. In order to investigate the effect of the method itself, we keep the number of particles and degrees of freedom the same for the model and the reference. Thus, the only "coarse graining" is in terms of swapping out the LJ potential with a tabulated IMC potential which we optimize. We derive effective interactions between all LJ particles with IMC at different number densities and compute the relative entropy[57] with respect to the reference LJ system at different mixture compositions. The relative entropy, which can be considered as a measure to assess how close a CG model is to a given target model[57, 58], shows a minimum for the derived models at a specific concentration. Correspondingly, the model derived at this concentration shows the best performance in terms of structural representability and transferability. This indicates the existence of an ideal concentration. Interestingly, all derived models show transferability in the direction of increasing concentration of the larger component. While most of the investigations are performed with the simplistic LJ system, we also progress to a more realistic system, i.e. n-hexane and n-perfluorohexane mixtures. We show that the trend in transferability of the pair structure can also be observed for an IMC model for these binary mixtures. The concentration transferable model for these systems is generated using IMC in combination with the previously proposed Dunn-Noid linear regression (DN-LR) scheme.[51]

The article is structured as follows: we first introduce the necessary theoretical background,

followed by the computational details of the simulations performed. Next, we present and discuss the main results of this study. We finally conclude this article by giving a short outlook on remaining challenges and open questions.

II. THEORETICAL BACKGROUND

A. Bottom-up coarse graining

Bottom-up coarse graining starts with the choice of the mapping scheme. This scheme determines the resolution of the CG model by assigning the FG particles, i, to effective interaction sites I, which are larger in size and mass. This assignment is determined by a mapping operator, $\mathbf{M}(\mathbf{r})$, which generates the CG configuration, $\mathbf{R}$, based on the positions of the FG particles, $\mathbf{r}$.[58]

$$\mathbf{R} = \mathbf{M}(\mathbf{r})) \tag{1}$$

Most commonly, one maps the coordinates of the particles ($\mathbf{r}_i$), which belong to the same CG particle, onto their center of mass to obtain the new position ($\mathbf{R}_I$).[58]

$$\mathbf{R}_I = \frac{\sum_{i\in I} m_i \mathbf{r}_i}{\sum_{i\in I} m_i} \tag{2}$$

Here, m_i is the mass of particle i which belongs to the CG bead I. As stated in the introduction, the mapping scheme should be selected with great care, despite the lack of a practicable method to do so.[36, 53–56]

On the basis of the selected mapping scheme, a new set of interaction potentials has to be generated for the CG particles. The optimal solution to this problem would be the so called multibody potential of mean force (m-PMF), which in reality is too complex to compute exactly. Thus, only approximate solutions can be determined.[6] To do so, we split the interaction potential, $U(\mathbf{R})$, into bonded ($U(r, \theta, \phi)$) and non-bonded ($U(r_{ij})$) contributions.

$$U(\mathbf{R}) = \sum_{i<j} U(r_{ij}) + \sum_{nbonds} U(r, \theta, \phi) \tag{3}$$

The first summation runs over all particle pairs, whereas the second sum is performed over the total number of bonds present in the system. Further, we decompose the bonded potentials into the potential energy which arises from the bond length (r) the bond angle (θ) and the torsional (ϕ) interaction.[59]

$$U(r, \theta, \phi) = U_r(r) + U_\theta(\theta) + U_\phi(\phi) \tag{4}$$

Each of those potential energies is the PMF ($U^0(q)$) obtained from the mapped probability distribution of the corresponding variable q, $P^0(q)$.[59]

$$U^0(q) = -k_B T \ln P^0(q) + C_q \tag{5}$$

where k_B is the Boltzmann constant, T the temperature of the system and C_q a constant. Here, the decomposition of interaction potentials (see Eqn. (4)) relies on the assumption that each contribution is decoupled from the others. The derivation of the non-bonded potential ($U(r_{ij})$) is explained in the following.

B. Inverse Monte Carlo

Non-bonded interactions are derived with the Inverse Monte Carlo (IMC) method, introduced by Lyubartsev and Laaksonen.[24] Here, the non-bonded interactions are determined to reproduce the structure of the mapped FG system in the CG configuration space. Thus, IMC is a structure based coarse graining technique. In IMC, the PMF is obtained from the mapped radial distribution function (RDF) between particles i and j. Opposite to the bonded interactions, where the PMF computed according to Eqn. 5 suffices, the non-bonded PMF ($U^{n-1}(r_{ij})$) has to be updated n- times with a term ΔU to accurately reproduce the reference structure. The iterative update is necessary since higher order correlations have to be mapped onto the pair wise additive potential for an accurate representability of the FG structure.

$$U^n(r_{ij}) = U^{n-1}(r_{ij}) + \Delta U(r_{ij}) \tag{6}$$

To determine ΔU a set of linear equations is iteratively solved until the structural difference between the CG and FG system is minimized.[24]

$$\mathbf{J}\Delta U_\alpha = \langle N_\alpha \rangle^{CG} - N_\alpha^{FG} \tag{7}$$

where $\langle N_\alpha \rangle$ is the average number of particle pairs at distance α either in the CG or FG system.[24] This number is related to the RDF ($g(r)$) via

$$N_\alpha = \frac{(N-1)}{2} \frac{4\pi r_\alpha^2 \Delta r}{V} g(r_\alpha) \tag{8}$$

where N is the total number of particles, V the bulk volume and $4\pi r_\alpha^2 \Delta r$ defines the size of local volume elements α. The **J**-matrix in Eq.(7) is the corresponding Jacobi matrix given by:[24]

$$\mathbf{J} = \frac{\partial N_\alpha}{\partial U_\gamma} = -\beta(\langle N_\alpha N_\gamma \rangle) - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle) \tag{9}$$

Here, β is $1/k_B T$.
IMC is an exact Newton inversion technique to solve the inverse problem of generating non-bonded pair potentials.[60] Therefore, the method leads to a better approximation of the m-PMF when compared to the similar iterative Boltzmann inversion method.[61]

C. Coarse graining in the isothermal, isobaric ensemble

CG models, which have been derived based on structure based coarse graining methods, often show weak cohesive energy between particles. This is due to the inaccurate representation of the long-ranged interactions.[23, 32, 33, 62] As a consequence the system expands in volume if a barostat is applied with the same pressure as during the FG simulations. The most common approach to tackle this problem is to add a linear ramp potential, which alters the tails of the effective pair potentials.[23, 33] Although widely spread, one of the major shortcomings of this method is that it changes the pair potentials after they have been parametrized to reproduce the structure at pair level. Thus, structural accuracy is lost.[33] To overcome this shortcoming Das and Andersen proposed another ansatz to enable CG simulations in the isothermal and isobaric (NPT) ensemble at the same pressure as for the FG system. They suggested to make the Hamiltonian of the CG system, $H(\mathbf{R}^N, \mathbf{p}^N, V)$, explicitly volume dependent.[46]

$$H(\mathbf{R}^N, \mathbf{p}^N, V) = \sum_{i=1}^{N} \frac{\mathbf{p_i}^2}{2m_i} + U(\mathbf{R}^N) + U_V(V) \tag{10}$$

In Eqn. (10), m_i is the mass, and $\mathbf{p_i}$ the corresponding momentum of particle i, and $U(\mathbf{R}^N)$ is the potential energy, which depends on the configuration of the CG system, $\mathbf{R}^N$. The additional volume dependent potential, $U_V(V)$, is independent of the configuration of the system and is defined as:[46]

$$U_V(V) = \frac{NV}{\langle V \rangle}\psi_1 + N\left(\frac{V}{\langle V \rangle} - 1\right)^2 \psi_2 \tag{11}$$

In Eqn. (11), N is the number of CG particles, V the volume of the system at time t, $\langle V \rangle$ is the average volume of the system obtained from the FG reference system and ψ_1 and ψ_2 are two unknown coefficients. In order to determine the two unknowns a variational principal is applied, the so called pressure matching. The goal is to minimize the pressure difference between the FG and CG model under the constraint of the previously derived CG potential energy, U_R.[46]

$$\chi^2(U_V|U_R) = \langle |P_{FG}(\mathbf{r}, \mathbf{p}, V) - (P^0_{CG}(\mathbf{M}(\mathbf{r}), \mathbf{M_p}(\mathbf{p})) + F_V(V))|^2 \rangle \tag{12}$$

In Eqn. (12), $P_{FG}(\mathbf{r}, \mathbf{p}, V)$ is the virial pressure of the FG system, $P^0_{CG}(\mathbf{M}(\mathbf{r}), \mathbf{M_p}(\mathbf{p}))$ is the pressure obtained when the CG potential is applied on the mapped FG configuration, $\mathbf{M}(\mathbf{r})$, including the mapped momenta, $\mathbf{M_p}(\mathbf{p})$, and $F_V(V)$ is:

$$F_V(V) = -\frac{\partial U_V(V)}{\partial V} \tag{13}$$

The average in Eqn. (12) is performed over the mapped FG configurations. If the CG model produces volume fluctuations different than those of the FG model, a self consistent correction of ψ_1 and ψ_2 has to be performed.[47, 48, 63] To lower the computational cost of the self consistent correction, we previously proposed to apply a linear regression scheme between two selected state points to predict the ψ values at different state points.[51]
To account for the volume as an additional degree of freedom, we add the volume dependent potential to the Hamiltonian derived by Martyna, Tobias, Tuckerman and Klein (MTTK) for MD simulations in the NPT ensemble.[64, 65]

$$H_1(V^{-1/3}\mathbf{R}^N, V^{1/3}\mathbf{p}_i, V, \Pi) = \sum_{i=1}^{N} \frac{V^{-1/3}\mathbf{p}_i^2}{2m_i V^{2/3}} + U(V^{1/3}, V^{-1/3}\mathbf{R}^N) + \frac{\Pi^2}{2M} + pV + U_V(V) \quad (14)$$

In Eqn. (14), V is the volume of the system, p is the pressure of the system, Π is the momentum of an imaginary piston acting on the system, M is the corresponding mass and all other variables have the same meaning as in Eqn. (10). The corresponding additional force term is implemented in the MTTK equations of motion and dampens the volume fluctuations regulated by the barostat. Thus, it compensates for the weaker cohesive energy between the CG particles.[46] Hence, the CG system can adjust its volume to predict the bulk density of a system at a given temperature, composition and pressure. This is a prerequisite for a transferable CG model.

III. SIMULATION DETAILS

A. Binary mixtures of Lennard-Jones particles

1. Reference simulations

To examine the existence of an ideal concentration for an IMC based CG model, we study a binary mixture of LJ particles LJ1 and LJ2, which are different in size. The particles interact via a 6-12 LJ potential.

$$U(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad (15)$$

Here, r_{ij} is the distance between two particles i and j, σ_{ij} is the minimum contact distance between two particles and ϵ is the potential well depth. The corresponding values for all interactions present in the system are given in table I. We examine the LJ systems at varying mole fractions x of the larger component LJ2. The total number of particles is always fixed at 500, and the side length of the cubic simulation box is kept constant at

TABLE I. Interaction parameter for the reference Lennard-Jones system

interaction type	σ_{ij} (nm)	ϵ_{ij} (kJ/mol)
LJ1-LJ1	0.296	0.84
LJ1-LJ2	0.317	0.91
LJ2-LJ2	0.340	0.98

TABLE II. Different mole fraction of LJ2 (x_{LJ2}), the corresponding particle numbers (N_{LJ1} and N_{LJ2}) and the reduced densities (ρ^*) for the systems studied.

x_{LJ2}	N_{LJ1}	N_{LJ2}	ρ^*
0.0	500	0	0.66
0.2	400	100	0.73
0.5	250	250	0.83
0.7	150	350	0.90
1.0	0	500	≈ 1.0

2.7 nm. The different compositions of the systems studied are presented in table II. MD simulations are performed using the stochastic dynamics integrator implemented in Gromacs (version 2016-4).[66, 67] The temperature is held constant at 85 K with an inverse friction constant of $\tau = 0.5$ ps $^{-1}$. The simulations are run for 50 ns with a time step of $\Delta t = 2fs$. Interactions are cut-off at 1.2 nm with a long-range dispersion correction for energy and pressure. Periodic boundary conditions are applied in x,y and z direction.

2. Generation of effective pair potentials

Based on the RDFs obtained from the reference simulations, we derive IMC models at three different mole fractions of LJ2 as presented in table III. The IMC models are generated

TABLE III. IMC models derived at different mole fraction of LJ2 (x_{LJ2}), the corresponding particle numbers (N_{LJ1} and N_{LJ2}) and the reduced densities ($\rho*$) for the systems studied and the number of iterations needed for convergence.

model	x_{LJ2}	N_{LJ1}	N_{LJ2}	$\rho*$	#iterations
IMC 0.2	0.2	400	100	0.73	30
IMC 0.5	0.5	250	250	0.83	60
IMC 0.7	0.7	150	350	0.90	30

with the VOTCA toolkit for coarse graining (developer version 1.5) [68, 69] in combination with Gromacs (version 2016-4).[66, 67] We perform 10 IMC iterations, which is enough to achieve convergence in structure and potential. The total number of iterations given in table III are larger as we had to perform some iterative Boltzmann inversion (IBI) simulations first. This was necessary to obtain a better initial guess than the simple PMF. Further, we applied a Tikhonov regularized version of IMC, which is implemented in VOTCA, with a regularization factor of 300. Details on that can be found elsewhere.[61] At each iteration, a short MD simulation of 2 ns is performed with the same settings as during the reference simulations. The cut-off value for the non-bonded interactions is set to 1.20 nm without any long range dispersion correction. The RDF between pairs of LJ1 particles is calculated between 0.28 nm and 1.20 nm, for pairs of LJ2 particles between 0.32 nm and 1.20 nm and for pairs of LJ1-LJ2 particles between 0.30 nm and 1.20 nm. The bin width for the RDF histogram is set to 0.01 nm in all cases. This applies to both, the IMC and IBI iterations. The performance of the three models is finally compared to the reference model at all different mixture compositions. Simulations are performed with the same setup as during the iterations with an increased simulation time of 50 ns.

B. Binary mixtures of n-hexane and n-perfluorohexane

1. *Fine grained simulations*

To test if our findings for the LJ system may also hold for actual CG models, we simulate binary mixtures (n)-hexane (HEX) and (n)-perfluorohexane (PFH). Here, PFH is the larger component and HEX the smaller one. Each system contains 500 molecules with varying mole fractions of PFH, $x_{\texttt{PFH}}$. As a FG force field we choose the TraPPE united atom model.[70] MD simulations are performed with the GROMACS simulation package (version 2016-4).[66, 67] Each system is equilibrated for 5 ns at a constant temperature of 300 K and at a constant pressure of 1 bar. Therefore, the Berendsen thermostat and barostat are applied with a coupling constant of τ_T=0.5 ps for the thermostat and τ_P=1.0 ps for the barostat.[71] The compressibility of the barostat is set to be $4.5 * 10^{-5}$ bar. Newton's equations of motion are integrated via a leap-frog algorithm with a timestep of 1 fs. The non-bonded interaction are cut-off at 1.4 nm and long range corrections for pressure and energy are applied. The bonded interaction are not constrained. For the final production run of 100 ns, the Nosé-Hoover thermostat[72] and the Parrinello-Rahman barostat[73] are applied with corresponding coupling constants of τ_T=0.5 ps and τ_P=5.0 ps. All other parameters are the same as during equilibration.

2. Coarse grained models

The CG models are based on a two to one mapping scheme, as illustrated in figure 1. Each molecule is built with two terminal beads of type A or C for HEX or PFH respectively, and one central bead of type B or D accordingly.

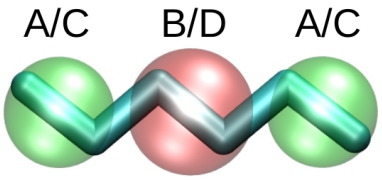


FIG. 1. Illustration of 3 bead mapping scheme for hexane (A-B-A) and perfluorohexane (C-D-C).

The bond and angular interactions for the two molecules are determined according to Eqn. (5). In order to obtain the required probability distributions, a single molecule is sampled in vacuum for 100 ns with a timestep of 1 fs. A Langevin dynamics thermostat with a inverse friction coefficient of 1 ps is applied to keep the temperature constant at 300 K.

The IMC potentials have been developed based on the target RDFs between the different bead types, which are obtained from liquid phase simulations of the FG system at $x_{\texttt{PFH}}$=0.5. To improve the convergence of the iterative procedure, we used pair potentials as an initial guess, which have been derived for this system with the conditional reversible work method in a recent study.[74] IMC calculations have been performed with the VOTCA toolkit (developer's version 1.5)[68, 69] in combination with GROMACS (version 2016-4).[66, 67] During each of the total 30 iterations short MD simulations of 2 ns are performed with a timestep of 1 fs. Simulations are executed at constant NVT conditions. A Langevin dynamics thermostat with a inverse friction coefficient of 1 ps guarantees a constant temperature of 300 K. The different RDFs between the CG beads are computed at different ranges as defined in table IV. For numerical stability of IMC, regions of poor sampling are removed from the RDF for pair distances smaller r_0. The cut-of values for the non-bonded interactions are chosen accordingly (r_{cut}). The grid spacing for all interactions is 0.01 nm.

3. Pressure matching

The two unknown coefficients in Eqn. (11), are determined by solving the variational problem defined in Eqn. (12). The solution is computed with a numpy script in python 2.7.[75] A more advanced version which is based on the same theoretical background is

TABLE IV. Starting (r_0) and end points (r_{cut}) for RDF evaluation between the different bead types during IMC.

interaction type	r_0 (nm)	r_{cut} (nm)
A-A	0.32	1.50
A-B	0.33	1.35
A-C	0.37	1.50
A-D	0.38	1.50
B-B	0.33	1.40
B-C	0.38	1.50
B-D	0.39	1.50
C-C	0.38	1.50
C-D	0.38	1.35
D-D	0.39	1.40

available within the BOCS package.[76] In order to obtain $P^0_{CG}(\mathbf{M}(\mathbf{r}), \mathbf{M}_\mathbf{p}(\mathbf{p}))$, a rerun of the mapped FG trajectory at $x_{\texttt{PFH}}$=0.5 is performed with the corresponding IMC potential. The necessary self consistent correction is performed at $x_{\texttt{PFH}}$=0.2 and $x_{\texttt{PFH}}$=0.5. MD simulations of 10 ns are performed with a modified version of the LAMMPS simulation package[77], which accounts for the additional volume dependent potential in the Hamiltonian (see Eqn. 14). During the simulations a time step of 1 fs is applied. The modified MTTK barostat[64, 65] with a coupling constant of τ_P=1.0 ps has been applied to ensure a constant pressure of 1 bar. Constant temperature is guaranteed by a Nosé-Hoover thermostat[72] with a chain length of 3 and a coupling constant of τ_T=0.2 ps. The self consistent approach is applied until no further change in the bulk density and the pressure is observed. The coefficients at mole fractions not included in the model parametrization are obtained via simple linear regression.[51]
The final CG simulations are performed for 100 ns at all mole fractions. The simulation parameter mentioned above were used in the production runs.

IV. RESULTS AND DISCUSSION

A. Binary mixtures of Lennard-Jones particles

Simulations of the binary LJ mixtures have been performed according to section III A 1. The computed RDFs for the target system are exemplary shown at $x_{\texttt{LJ2}} = 0.5$ in figure 2.

As one sees, the RDFs reflect the trend of the different LJ parameters given in table I.

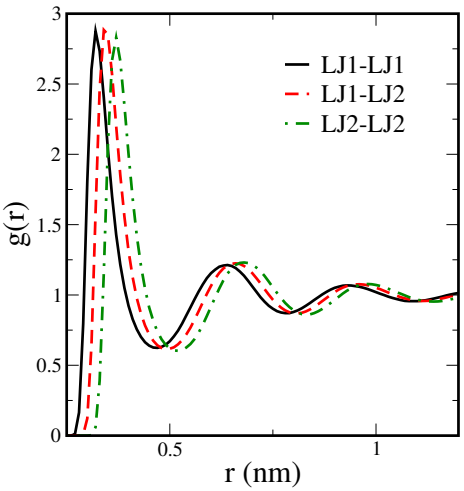


FIG. 2. Radial distribution function between all three LJ components at $x_{\mathrm{LJ2}} = 0.5$: LJ1-LJ1 solid black line, LJ1-LJ2 dashed red line and LJ2-LJ2 dashed-dotted green line.

According to the procedure described in section III A 2, three IMC models have been derived at $x_{\mathrm{LJ2}} = 0.2$, $x_{\mathrm{LJ2}} = 0.5$ and $x_{\mathrm{LJ2}} = 0.7$ based on the corresponding RDFs for all three interactions. The complementary models are called IMC 0.2, IMC 0.5 and IMC 0.7 respectively. The final potentials are shown in figure 3 in comparison with the reference LJ potential (FG), which is independent of concentration and which was used to generate the target distributions at the corresponding concentrations for the different models.

In all graphs in figure 3 one sees that the IMC 0.5 model (dotted-dashed green line) and the IMC 0.7 model (dashed-dotted-dotted blue line) can capture the potential well of the reference LJ potential (solid black line) for all three interactions LJ1-LJ1 (figure 3 (a)), LJ1-LJ2 (figure 3 (b)) and LJ2-LJ2 (figure 3 (c)). But, at a distance of ≈ 0.6 nm a second minimum starts to evolve, which is absent in the LJ potential. This shows that the IMC potentials are not able to capture the tail behavior of a LJ potential, which is in agreement with the literature.[61, 78] The IMC 0.2 model (dashed red line) is more repulsive than the other two IMC models. It neither captures the potential well nor the correct tail behavior for all three interactions. Despite having different potential shapes, all IMC models reproduce the target RDFs at the concentration chosen for parametrization, as shown for the LJ1-LJ1 interaction in figure 5.

In order to quantify the differences between the IMC models and the LJ model, we compute the relative entropy (S_{rel}) at different x_{LJ2}. The relative entropy measures the overlap between the probability distribution q of configurations i in the model (M) and the underlying target system (T). It can be derived based on the Kullback-Leibler divergence[79] and

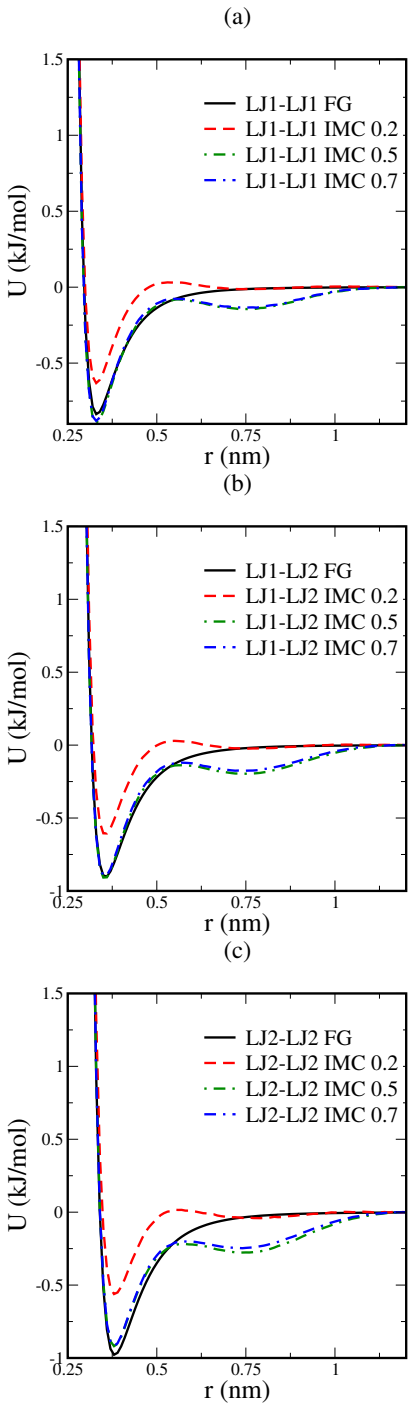


FIG. 3. Final IMC potentials for the LJ1-LJ1 (a), the LJ1-LJ2 (b) and the LJ2-LJ2 (c) interaction for the three different models, IMC 0.2 (dashed red line), IMC 0.5 model (dotted-dashed green line) and IMC 0.7 model (dashed-dotted-dotted blue line) in comparison with the reference LJ potential (solid black line).

computes as follows:[57]

$$S_{rel} = \sum_i q_i^T \ln \frac{q_i^T}{q_i^M} \tag{16}$$

If $q_i^T = q_i^M$, S_{rel} equals zero, i.e. the smaller S_{rel} the larger the overlap between the probability distributions of target and model and the more accurate is the model. In the canonical ensemble S_{rel} can be calculated according to:[57]

$$S_{rel} = \beta\langle U^M - U^T\rangle_T - \beta(A^M - A^T) \tag{17}$$

$\langle U^M\rangle_T$ is the potential energy when the IMC potential is applied on a target configuration generated with the LJ potential, whereas $\langle U^T\rangle_T$ is the potential energy of the reference system. The configurational part of the Helmholtz free energy (A) is computed by transferring the target and the model system from an ideal gas state to a system with all interactions turned on at constant NVT conditions corresponding to the number densities given in table II. Here, free energies are evaluated with thermodynamic integration (TI)[80] and Bennett's acceptance ratio (BAR) method.[81] We used 20 equally distributed λ points between 0 (ideal gas) and 1 (real system). We applied a soft core potential (V_{sc}) for the intermediate λ states.

$$V_{sc} = (1-\lambda)V_a(r(\lambda)) + \lambda V_b(r(\lambda)) \tag{18}$$

where V_a is the interaction potential for the ideal gas system, and V_b for the real system. $r(\lambda)$ is a shifted distance between two particles i and j (r_{ij}), which depends on $\sigma = 0.3$ nm (in this study) and λ. It is defined as.

$$r(\lambda) = (\sigma_b^6(1-\lambda) + r_{ij}^6)^{1/6} \tag{19}$$

The simulations were run for 10 ns, with the same setup as for the reference/IMC simulations.

In figure 4 (a) the free energy changes ΔA as a function of λ are exemplified for $x_{\mathtt{LJ2}} = 0.5$. The depicted error bars correspond the standard deviation. The results shown are obtained from BAR, the corresponding TI results (not shown here) are comparable except for the last two λ points. There, the derivatives of the tabulated IMC potentials with respect to λ start to diverge, which is most likely related to numerical issues as no irreversible effects as phase transitions are observed. The free energy change for the IMC 0.2 model (red squares) is very small. It seems that the initial ideal gas state is close to the equilibrium state of the system with all interactions turned on. The IMC 0.5 (green diamonds) and IMC 0.7 model (blue upper triangles) show a decrease in the free energies as the interactions are turned on. This follows the trend of the FG model (black circles). But, the CG models do not reach a well defined plateau as the FG model does. The final ΔA values ($\lambda = 1.0$) are taken to

compute the relative entropy.

In figure 4 (b) the relative entropy values per molecule are shown as a function of x_{LJ2}. We only present data for the IMC 0.5 model (green diamonds) and IMC 0.7 model (blue upper triangles), as the IMC 0.2 model gives negative relative entropies. This is not possible per definition, as the relative entropy is a strictly positive quantity.[57]. The negative relative entropy arises from the small change in the free energy as the systems gets transferred from an ideal gas to the real system. The lack of a decrease in ΔA can either be a consequence of an improper sampling of the configuration space caused by the repulsive nature of the IMC 0.2 pair potentials or by an improper $\lambda-$path to compute the free energies. Nevertheless one see that the IMC 0.5 model and IMC 0.7 model both have the lowest relative entropy at $x_{\mathrm{LJ2}} = 0.2$. This is surprising as one would expect the lowest relative entropy to appear at the concentration chosen for parametrization. What is more, the IMC 0.7 model has a lower relative entropy than the IMC 0.5 model.

In order to examine the practical consequences of the different relative entropies, we investigate the structural transferability of the three IMC models. For convenience, we discuss the transferability based on the LJ1-LJ1 interaction in the following. Results and their discussion can be extended to the remaining interactions as well, since they show alike behavior. In figure 5, one sees that the IMC 0.2 model (dashed red line) reproduces the RDF of the LJ system not only at its reference point of $x_{\mathrm{LJ2}} = 0.2$ (figure 5 (a)), but also at $x_{\mathrm{LJ2}} = 0.5$ (figure 5 (b)) and $x_{\mathrm{LJ2}} = 0.7$ (figure 5 (c)). The IMC 0.5 model (dotted-dashed green line) and the IMC 0.7 model (dashed-dotted-dotted blue line) can only reproduce the RDFs at mole fractions of $x_{\mathrm{LJ2}} = 0.5$ and $x_{\mathrm{LJ2}} = 0.7$. Further, if we go to the pure systems, i.e. $x_{\mathrm{LJ2}} = 0.0$ (figure 6 (a)) and $x_{\mathrm{LJ2}} = 1.0$ (figure 6 (b)), again the IMC 0.2 model (dashed red line) can reproduce the RDF of the LJ model (solid black line). Instead, the IMC 0.5 model (dotted-dashed green line) and the IMC 0.7 model (dashed-dotted-dotted blue line) can only reproduce the structure at $x_{\mathrm{LJ2}} = 1.0$, where the system is in a solid state.

To conclude, all IMC models show transferability for the RDF in the direction of increasing concentration of the larger component LJ2. Further, the IMC 0.7 model shows also transferability down to $x_{\mathrm{LJ2}} = 0.5$. This transferability in both directions also corresponds to the lower relative entropy of the IMC 0.7 model compared to the IMC 0.5 model, which is only transferable in the direction of increasing x_{LJ2}. The IMC 0.2 model, which has been derived at a concentration where both other models have the lowest value in the relative entropy, shows transferability over the whole concentration range investigated. This means, it also captures a phase transition from liquid to solid as the concentration of LJ2 increases. But, it does not mean that $x_{\mathrm{LJ2}} = 0.2$ is a true optimum in the concentration space for this specific system. It is rather an ideal concentration , since it remains unclear if there is just a single point or a range with a boundary value in the concentration space at which the derived

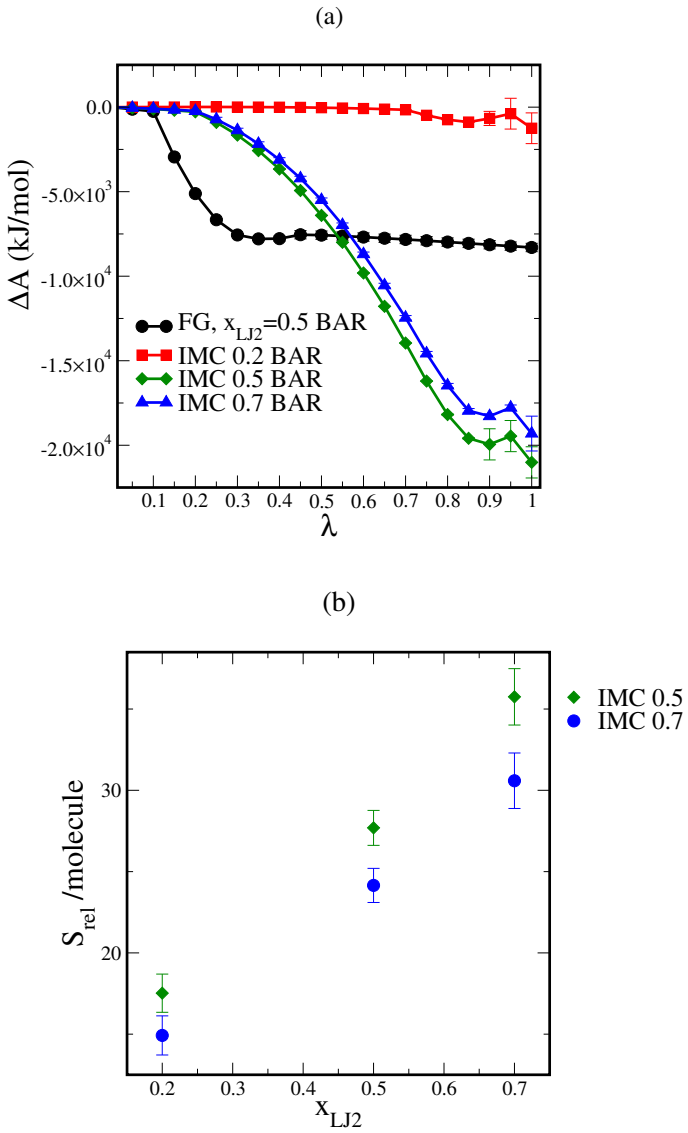


FIG. 4. (a): free energy changes as a function of λ at x_{LJ2} =0.5 for the FG system (black circles) and the IMC 0.2 model (red squares), the IMC 0.5 model (green diamonds) and the IMC 0.7 model (blue uper triangles); (b): relative entropy per molecule between the FG system the IMC 0.5 model (green diamonds) and the IMC 0.7 model (blue upper triangles) as function of $x_{\mathtt{LJ2}}$

potentials show the same behavior in terms of transferability. Nevertheless, we show that it is possible to overcome problems in terms of structural transferability by systematically selecting a concentration to parametrize a CG model for binary mixtures.

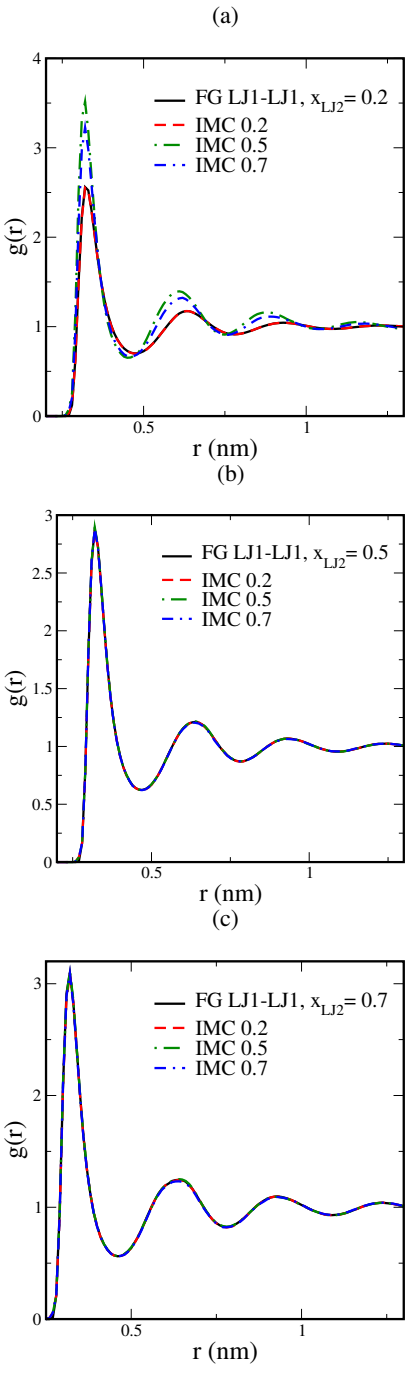


FIG. 5. Radial distribution function between LJ1 particles obtained from the FG (solid black line), the IMC 0.2 (dashed red line), the IMC 0.5 (dotted-dashed green line) and the IMC 0.7 (dashed-dotted-dotted blue line) model at $x_{\mathrm{LJ2}} = 0.2$ (a), $x_{\mathrm{LJ2}} = 0.5$ (b) and $x_{\mathrm{LJ2}} = 0.7$ (c).

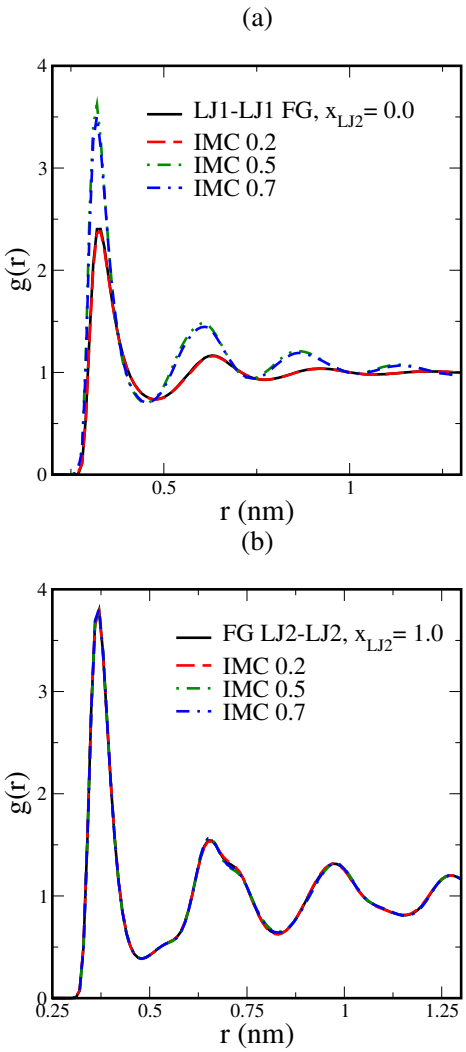


FIG. 6. Radial distribution function between LJ1 particles obtained from the FG(solid black line), the IMC 0.2 (dashed red line), the IMC 0.5 (dotted-dashed green line and the IMC 0.7 (dashed-dotted-dotted blue line) model at $x_{\mathrm{LJ2}} = 0.0$ (a) and $x_{\mathrm{LJ2}} = 1.0$ (b).

B. Binary mixtures of n-hexane and n-perfluorohexane

1. *Coarse grained potentials*

So far we have focused on binary mixtures of LJ particles. In the following, we want to investigate to what extent the main conclusions of the previous section hold for a binary mixture of different sized real, chemical compounds. Therefore, we have derived a CG model for mixtures of HEX and PFH with IMC according to section III B 2.

The bonded potentials for CG HEX and PFH are obtained via simple Boltzmann inversion

(see Eqn. (5)) of the probability distribution of the bond length between A-B and C-D CG sites and of the A-B-A and C-D-C bond angle. As shown in figure 7, the potentials for the bond distance (see figure 7 (a)) and the bond angle (see figure 7 (b)) differ significantly between HEX (solid black curves) and PFH (dashed red curves). This difference reflects the discrepancy in the underlying probability distributions of the bond length and the bond angle for the two molecules caused by the distinct dihedral interactions present in the underlying FG model. The two minima in each of the bond length potentials correspond to the gauché and trans state of the HEX/PFH chains. Both bond angle potentials show a minimum for the trans state (180°). This indicates, that the cis state of the chains is not sampled with the derived CG models. Based on the assumption that the bonded interac-

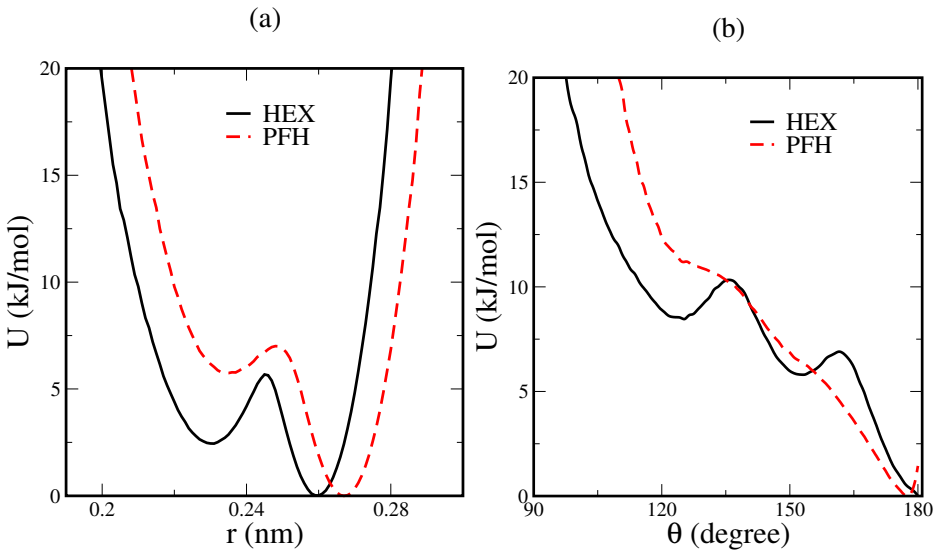


FIG. 7. (a): Bond length potential for HEX (solid black curve) and PFH (dashed red curve) obtained via Boltzmann inversion; (b): Bond angle potential for HEX (solid black curve) and PFH (dashed red curve) obtained via Boltzmann inversion.

tions are decoupled from the non-bonded interactions, the latter ones have been derived with IMC. Further, as the beads are uncharged there are no long range interactions present in the system.

In figure 8, the non-bonded IMC potentials are presented together with the corresponding RDFs of the CG and FG reference system. As one sees, all reference site-site RDFs are accurately reproduced by the CG models, indicating the convergence of the iterative scheme. Further, all interactions between terminal beads, A-A, A-C and C-C, are more attractive than those involving the central B/D beads. These interactions also show a small potential barrier around 0.7 nm, which indicates that direct interactions with the central B/D sites

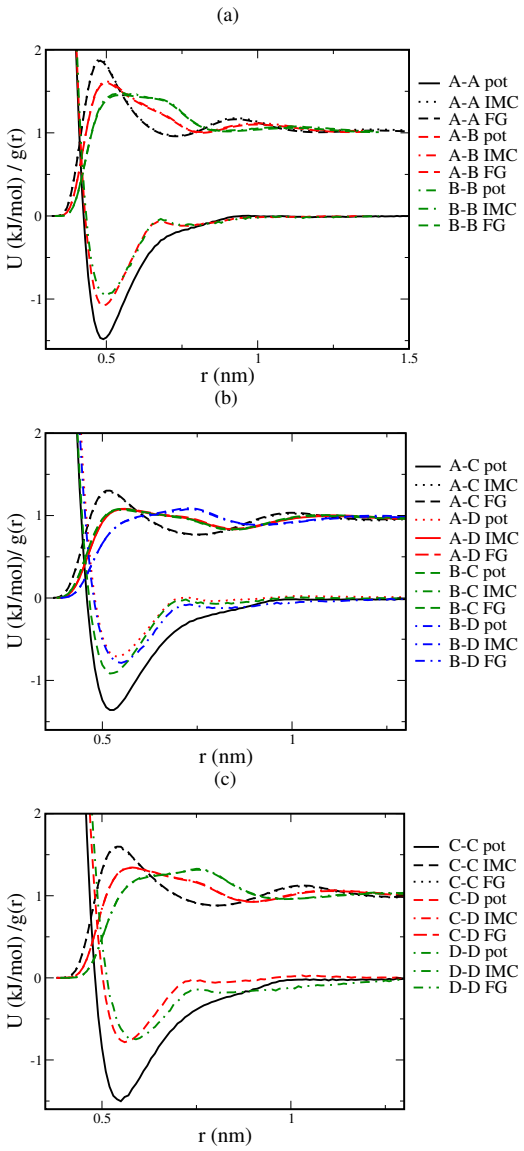


FIG. 8. Non-bonded IMC potentials together with the corresponding RDFs of the underlying FG system and the CG model; (a): site-site potentials of HEX; (b): intra-special site-site potentials; (c): site-site potentials of PFH.

are sterically constraint by the larger terminal A/C sites.

To further assess the quality of the different non-bonded potentials, we perform NVT simulations at $x_{\mathtt{PFH}} = 0.5$ and 300 K and computed the center of mass (com) RDFs, between HEX-HEX, HEX-PFH and PFH-PFH. As one sees in figure 5, the IMC model matches all com RDFs obtained from the FG system very accurately. The com RDFs also reflect the larger size of PFH compared to HEX.

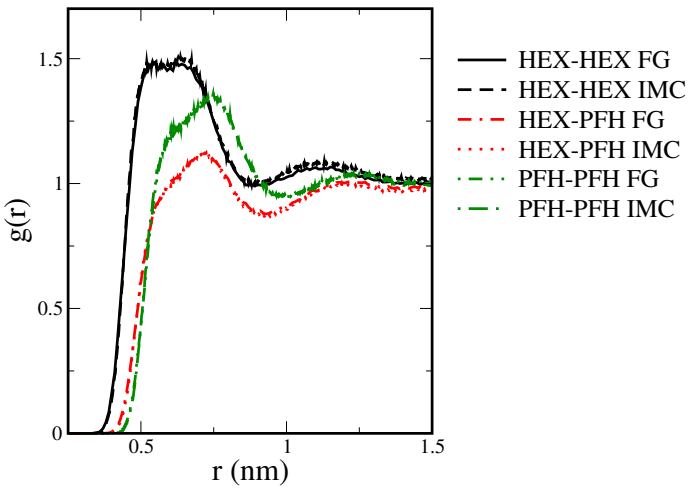


FIG. 9. Center of mass RDFs between HEX-HEX (solid black curves for FG model, dashed black curve for IMC model), HEX-PFH (dashed-dotted red curve for FG model, dotted red curve for IMC model) and PFH-PFH (dashed-dotted-dotted green curve for FG model, dotted-dashed green curve for IMC model).

In order to enable CG simulations with the correct bulk density at 1 bar, the additional volume dependent potential $U_V(V)$ (see Eqn. (10)) has to be determined. Therefore, we perform the pressure matching at $x_{\mathtt{PFH}}$=0.5. The set of coefficients obtained are shown in table V and are called Das-Andersen (DA) from now on. Based on the DA coefficients, the self consistent correction proposed by Dunn and Noid (DN)[47] is performed at $x_{\mathtt{PFH}}$=0.2 and $x_{\mathtt{PFH}}$=0.5. The optimized values for ψ_1 and ψ_2 (DN coefficients) are used for the Dunn Noid-Linear Regression (DN-LR) approach in order to predict the coefficients at different mole fractions. All values are presented in table V. As one sees, the values for ψ_1 increase with increasing $x_{\mathtt{PFH}}$, whereas ψ_2 decreases with increasing $x_{\mathtt{PFH}}$.

2. Concentration transferability

After determining the coefficients for the volume dependent potential, CG simulations are performed under NPT conditions at 1 bar with (IMC DN-LR) and without (IMC) the volume dependent potential. In order to evaluate the effect of $U_V(V)$, we compute the bulk densities and the isothermal compressibility as a function of $x_{\mathtt{PFH}}$. All results are shown in figure 10.

We show the bulk densities as produced by the TraPPE-united atom model (black circles) in comparison to the IMC DN-LR model (red diamonds) and the IMC model (green squares) in figure 10 (a). As one sees, the IMC DN-LR model accurately reproduces the bulk densities

TABLE V. Coefficients ψ_1 and ψ_2 obtained from pressure matching (DA), the self-consistent correction (DN) and the linear regression approach (DN-LR).

CG model	$x_{\mathtt{PFH}}$	method	ψ_1 (bar nm^3)	ψ_2 (bar nm^3)
IMC	0.0	DN-LR	9.767	71.067
	0.1	DN-LR	10.623	59.015
	0.2	DA	-33.964	41.632
	0.2	DN	11.479	46.944
	0.5	DA	-33.964	41.632
	0.5	DN	14.046	10.760
	0.7	DN-LR	15.757	-13.360
	1.0	DN-LR	18.324	-49.542

almost over the whole concentration range investigated, with an exception at $x_{\mathtt{PFH}} = 1.0$. Here, the density of the underlying FG model gets slightly overestimated. Contrary to that, the IMC model underestimates the bulk densities, as shown at $x_{\mathtt{PFH}} = 0.0$, $x_{\mathtt{PFH}} = 0.5$ and $x_{\mathtt{PFH}} = 1.0$. This shows the necessity of $U_V(V)$ in order to ensure correct bulk density if a barostat of 1 bar is applied during CG simulations. Further, it shows the transferability of the IMC potentials to different concentrations and the validity of the DN-LR approach to predict $U_V(V)$. For the isothermal compressibility (κ_T) a similar trend can be found, as shown in figure 10 (b). If $U_V(V)$ is absent in the model, the IMC potentials (green squares) produce a higher compressibility compared to the TraPPE reference model (black circles). This corresponds to the the lower bulk densities, which the IMC model produces. Again, the IMC DN-LR model (red diamonds) shows improved representability and transferability now in terms of the isothermal compressibility. This emphasizes the transferability of the IMC DN-LR model in terms of thermodynamic properties.

As we have applied IMC to generate the non-bonded interactions, we also want to assess the structural transferability of the derived CG model. Therefore, we compute the RDFs between the com of HEX-HEX at $x_{\mathtt{PFH}} = 0.0$ and PFH-PFH at $x_{\mathtt{PFH}} = 1.0$ based on the IMC DN-LR models. In figure 11 (a), one sees that the IMC DN-LR model (dashed black curves for HEX and dotted-dashed red curves for PFH) accurately represents the RDFs of the underlying FG systems (solid black curve for HEX and red dotted curve for PFH), although no information of these concentrations has been included during parametrization. Noticeably, it seems that for the pure PFH case the structural accuracy is slightly better compared to the pure HEX case. This is further indicated by the root mean square (RMS) error in the com RDF between the FG and CG system for HEX-HEX (solid black curve in

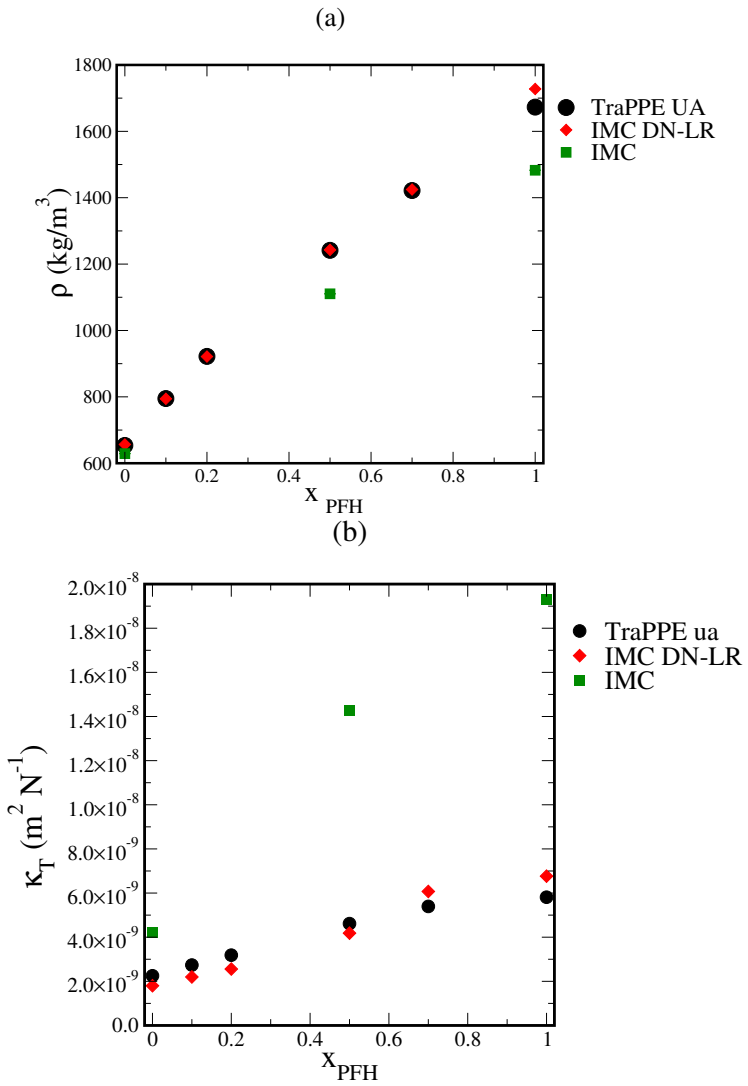


FIG. 10. Results of the CG simulations (IMC-DNLR: red diamonds; IMC: green squares) at 1 bar in comparison with the FG reference (black circles); (a): bulk densities ρ ; (b): isothermal compressibility κ_T.

figure 11 (b)) and for PFH-PFH (dashed red curve in figure 11 (b)). The `RMS error` can be cast as:

$$\texttt{RMS error} = \sqrt{\frac{\sum_i^N (g^{FG}(r) - g_i^{CG}(r))^2}{N}}, \tag{20}$$

where the summation is performed over $N = 5$ CG simulations.

The slightly smaller `RMS error` for the pure PFH case indicates that the IMC model is more transferable towards the direction of increasing concentration of the larger component PFH, which is in agreement to the conclusion drawn from the simple LJ system. Further, this

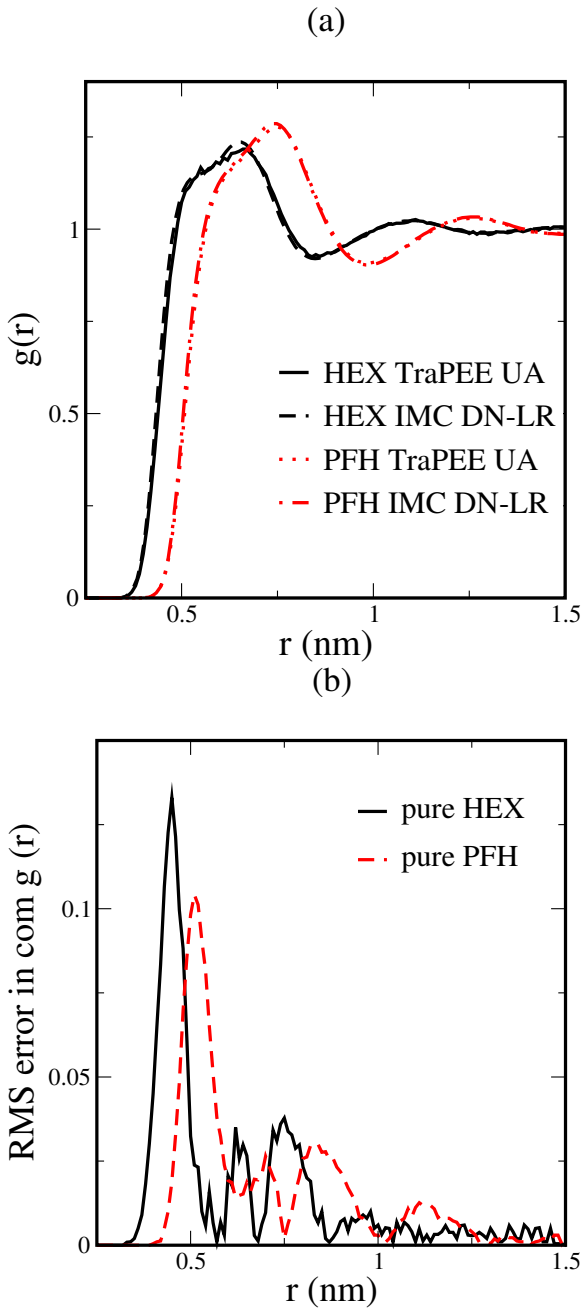


FIG. 11. (a): Center of mass (com) RDF between HEX-HEX (solid black curve FG model, dashed black curve IMC-DNLR model) and PFH-PFH (dotted red curve FG model, dotted-dashed red curve IMC-DNLR model); (b): RMS error in the com g(r) between the FG and CG system (solid black line HEX; dashed red line PFH).

is contrary to what one might conclude from the transferability in terms of the bulk density and the isothermal compressibility. There, it seems that the model is more transferable in the direction of increasing concentration of the smaller component HEX. This illustrates one of the major shortcomings of structure-based coarse graining that matching structures does not equal matching thermodynamics. But with the additional volume dependent potential term a good compromise can be achieved.

The reason we do not show any relative entropies for the HEX/PFH system is owed to the fact that computing relative entropies for real CG systems underlies several constraints. First, in terms of coarse graining the relative entropy will get an additional contribution, the so called mapping entropy. This contribution accounts for the degeneracy of states in the CG model and cannot be straightforwardly computed, but also cannot be ignored.[54, 58] Thus, one cannot disentangle effects arising from mapping and arising from the applied coarse graining method. Second, the bonded and angular potentials have also an effect on the computed energies. Here, the IMC models are derived under the assumptions that bonded and non-bonded potentials are fully decoupled. Although this assumption seems to be valid, as we can reproduce several structural and thermodynamic properties, it requires further investigations of the effect of bonded potentials on the sampling of the configuration space. Finally, the free energy calculations might suffer from some inconsistencies, as we have to deal with molecular ideal gases with different degrees of freedom as reference states.

V. CONCLUSION AND OUTLOOK

We addressed the question if it is possible to overcome methodological disadvantages of IMC in terms of representability and transferability of the derived effective pair potentials by systematically selecting a state point for the model parametrization. With the help of the relative entropy we could identify such a concentration for binary mixtures of LJ particles different in size. Models parametrized at this ideal concentration are structurally transferable over the whole concentration range investigated. This does not automatically guarantee that this state point also corresponds to an optimum in the concentration space. It is possible that there is a range of concentrations at which all derived models show the same behavior. Nevertheless, all generated models show structural transferability in the direction of increasing concentration of the larger component regardless of their underlying reference point. This trend is also observed for a CG model of HEX/PFH, although to a smaller extent. This suggests that for binary mixtures of different sized particles, a structure based coarse graining model can rather be developed at higher concentrations of the smaller component. An aspect one should keep in mind that transferability of a bottom-up CG model cannot be simply assumed, one always has to validate the model against some refer-

ence data. Nevertheless, we show that it is possible to achieve transferability by systematic considerations.
Despite being able to predict the direction of transferability, we also showed that reproducing the pair structure might not be a useful target to ensure consistent sampling between the CG and FG model. This can have severe consequences on representability and transferability of other thermodynamic properties. Further, this clearly shows that methodological aspects cannot be fully ignored and that the search of ideal concentration might always apply for a specific target property only. Thus, it would be interesting to investigate if similar observations can be made for other coarse graining methods like force matching[20, 21] or relative entropy optimization[57] and for other temperatures or pressures.
Additionally, actual computation of relative entropies for different CG systems is inevitable to search for an ideal state point to parametrize a CG model for a given system. Hence, the effect of mapping and bonded interactions on the relative entropy have to be studied in greater detail to claim that such an ideal state point might exist for any given system. Following this lines, the simple brute force approach applied in this study might be computationally not feasible to do so. It might be worth to consider a combination with manifold learning techniques in order to speed up the exploration of a given phase space.[82, 83]
Regardless of the presented shortcomings or uncertainties, we think this study adds a new flavor to the field of bottom-up coarse graining, as we could show that shortcomings of a given method can be minimized by carefully selecting a reference point for parametrization. Consequently, we see this work as a starting point worth to explore more in the future.

ACKNOWLEDGEMENT

The authors thank Marvin Bernhardt, Swaminath Bharadwaj and Tanmoy Sanyal for critical discussions and helpful comments on the manuscript. Financial support is granted by the Deutsche Forschungsgmeinschaft (DFG, German Research Foundation) through Project number 233630050- TRR 146.

[1] F. Müller-Plathe, "Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back," *ChemPhysChem*, vol. 3, pp. 754–769, Sept. 2002.

[2] C. Peter and K. Kremer, "Multiscale simulation of soft matter systems from the atomistic to the coarse-grained level and back," *Soft Matter*, vol. 5, no. 22, p. 4357, 2009.

[3] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, "Multiscale modeling of emergent materials: biological and soft matter," *Physical Chemistry Chemical Physics*,

vol. 11, no. 12, p. 1869, 2009.

[4] C. Peter and K. Kremer, "Multiscale simulation of soft matter systems," *Faraday Discuss.*, vol. 144, pp. 9–24, 2010.

[5] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt, "Systematic coarse-graining methods for soft matter simulations a review," *Soft Matter*, vol. 9, no. 7, pp. 2108–2119, 2013.

[6] W. G. Noid, "Perspective: Coarse-grained models for biomolecular systems," *The Journal of Chemical Physics*, vol. 139, p. 090901, Sept. 2013.

[7] R. Potestio, C. Peter, and K. Kremer, "Computer Simulations of Soft Matter: Linking the Scales," *Entropy*, vol. 16, pp. 4199–4245, July 2014.

[8] M. Deserno, "Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications," *Macromolecular Rapid Communications*, vol. 30, pp. 752–771, May 2009.

[9] F. Schmid, "Toy amphiphiles on the computer: What can we learn from generic models?," *Macromolecular Rapid Communications*, vol. 30, pp. 741–751, May 2009.

[10] J. D. Honeycutt and D. Thirumalai, "Metastability of the folded states of globular proteins," *Proc Natl Acad Sci USA*, vol. 87, p. 3526, May 1990.

[11] I. R. Cooke, K. Kremer, and M. Deserno, "Tunable generic model for fluid bilayer membranes," *Physical Review E*, vol. 72, July 2005.

[12] E. Brini, V. Marcon, and N. F. A. van der Vegt, "Conditional reversible work method for molecular coarse graining applications," *Physical Chemistry Chemical Physics*, vol. 13, no. 22, p. 10468, 2011.

[13] Y. Wang, W. G. Noid, P. Liu, and G. A. Voth, "Effective force coarse-graining," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, p. 2002, 2009.

[14] N. Zacharopoulos, N. Vergadou, and D. N. Theodorou, "Coarse graining using pretabulated potentials: Liquid benzene," *The Journal of Chemical Physics*, vol. 122, p. 244111, June 2005.

[15] B. Hess, C. Holm, and N. F. A. van der Vegt, "Modeling Multibody Effects in Ionic Solutions with a Concentration Dependent Dielectric Permittivity," *Physical Review Letters*, vol. 96, Apr. 2006.

[16] B. M. Mognetti, L. Yelash, P. Virnau, W. Paul, K. Binder, M. Müller, and L. G. MacDowell, "Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide," *The Journal of Chemical Physics*, vol. 128, p. 104501, Mar. 2008.

[17] A. Villa, C. Peter, and N. F. A. van der Vegt, "Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water," *Journal of Chemical Theory and Computation*, vol. 6, pp. 2434–2444, Aug. 2010.

[18] F. Ercolessi and J. B. Adams, "Interatomic Potentials from First-Principles Calculations: The Force-Matching Method," *Europhysics Letters (EPL)*, vol. 26, pp. 583–588, June 1994.

[19] S. Izvekov and G. A. Voth, "A Multiscale Coarse-Graining Method for Biomolecular Systems," *The Journal of Physical Chemistry B*, vol. 109, pp. 2469–2473, Feb. 2005.

[20] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *The Journal of Chemical Physics*, vol. 128, p. 244114, June 2008.

[21] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, "The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models," *The Journal of Chemical Physics*, vol. 128, p. 244115, June 2008.

[22] W. Schommers, "Pair potentials in disordered many-particle systems: A study for liquid gallium," *Physical Review A*, vol. 28, pp. 3599–3605, Dec. 1983.

[23] D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations: Mesoscale Potentials from Atomistic Simulations," *Journal of Computational Chemistry*, vol. 24, pp. 1624–1636, Oct. 2003.

[24] A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," *Physical Review E*, vol. 52, pp. 3730–3737, Oct. 1995.

[25] J. W. Mullinax and W. G. Noid, "Generalized Yvon-Born-Green Theory for Molecular Systems," *Physical Review Letters*, vol. 103, Nov. 2009.

[26] J. W. Mullinax and W. G. Noid, "A Generalized-YvonBornGreen Theory for Determining Coarse-Grained Interaction Potentials ," *The Journal of Physical Chemistry C*, vol. 114, pp. 5661–5674, Apr. 2010.

[27] J. F. Rudzinski and W. G. Noid, "Investigation of Coarse-Grained Mappings via an Iterative Generalized YvonBornGreen Method," *The Journal of Physical Chemistry B*, vol. 118, pp. 8295–8312, July 2014.

[28] J. F. Rudzinski and W. G. Noid, "A generalized-Yvon-Born-Green method for coarse-grained modeling: Advances, Challenges, and Insight," *The European Physical Journal Special Topics*, vol. 224, pp. 2193–2216, Sept. 2015.

[29] J. Ghosh and R. Faller, "State point dependence of systematically coarse-grained potentials," *Molecular Simulation*, vol. 33, pp. 759–767, Aug. 2007.

[30] E. C. Allen and G. C. Rutledge, "A novel algorithm for creating coarse-grained, density dependent implicit solvent models," *The Journal of Chemical Physics*, vol. 128, p. 154115, Apr. 2008.

[31] P. Ganguly and N. F. A. van der Vegt, "Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations," *Journal of Chemical Theory and Computation*, vol. 9, pp. 1347–1355, Mar. 2013.

[32] M. E. Johnson, T. Head-Gordon, and A. A. Louis, "Representability problems for coarse-grained water potentials," *The Journal of Chemical Physics*, vol. 126, p. 144509, Apr. 2007.

[33] H. Wang, C. Junghans, and K. Kremer, "Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining?," *The European Physical Journal E*, vol. 28, pp. 221–229, Feb. 2009.

[34] S. Riniker, J. R. Allison, and W. F. van Gunsteren, "On developing coarse-grained models for biomolecular simulation: a review," *Physical Chemistry Chemical Physics*, vol. 14, no. 36, p. 12423, 2012.

[35] M. Guenza, "Thermodynamic consistency and other challenges in coarse-graining models," *The European Physical Journal Special Topics*, vol. 224, pp. 2177–2191, Sept. 2015.

[36] J. F. Rudzinski and W. G. Noid, "Bottom-Up Coarse-Graining of Peptide Ensembles and HelixCoil Transitions," *Journal of Chemical Theory and Computation*, vol. 11, pp. 1278–1291, Mar. 2015.

[37] D. Fritz, V. A. Harmandaris, K. Kremer, and N. F. A. van der Vegt, "Coarse-Grained Polymer Melts Based on Isolated Atomistic Chains: Simulation of Polystyrene of Different Tacticities," *Macromolecules*, vol. 42, pp. 7579–7588, Oct. 2009.

[38] E. Brini and N. F. A. van der Vegt, "Chemically transferable coarse-grained potentials from conditional reversible work calculations," *The Journal of Chemical Physics*, vol. 137, p. 154113, Oct. 2012.

[39] E. Brini, C. R. Herbers, G. Deichmann, and N. F. A. van der Vegt, "Thermodynamic transferability of coarse-grained potentials for polymeradditive systems," *Physical Chemistry Chemical Physics*, vol. 14, no. 34, p. 11896, 2012.

[40] T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen, "Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints," *The Journal of Chemical Physics*, vol. 126, p. 075101, Feb. 2007.

[41] P. Ganguly, D. Mukherji, C. Junghans, and N. F. A. van der Vegt, "KirkwoodBuff Coarse-Grained Force Fields for Aqueous Solutions," *Journal of Chemical Theory and Computation*, vol. 8, pp. 1802–1807, May 2012.

[42] A. Moradzadeh, M. H. Motevaselian, S. Y. Mashayak, and N. R. Aluru, "Coarse-Grained Force Field for Imidazolium-Based Ionic Liquids," *Journal of Chemical Theory and Computation*, vol. 14, pp. 3252–3261, June 2018.

[43] M. H. Motevaselian, S. Y. Mashayak, and N. R. Aluru, "Extended coarse-grained dipole model for polar liquids: Application to bulk and confined water," *Physical Review E*, vol. 98, Nov. 2018.

[44] J. W. Mullinax and W. G. Noid, "Extended ensemble approach for deriving transferable coarse-grained potentials," *The Journal of Chemical Physics*, vol. 131, no. 10, p. 104110, 2009.

[45] T. C. Moore, C. R. Iacovella, and C. McCabe, "Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion," *The Journal of Chemical Physics*, vol. 140, p. 224104, June 2014.

[46] A. Das and H. C. Andersen, "The multiscale coarse-graining method. V. Isothermal-isobaric ensemble," *The Journal of Chemical Physics*, vol. 132, p. 164106, Apr. 2010.

[47] N. J. H. Dunn and W. G. Noid, "Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids," *The Journal of Chemical Physics*, vol. 143, p. 243148, Dec. 2015.

[48] N. J. H. Dunn and W. G. Noid, "Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures," *The Journal of Chemical Physics*, vol. 144, p. 204124, May 2016.

[49] T. Sanyal and M. S. Shell, "Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation," *The Journal of Chemical Physics*, vol. 145, p. 034109, July 2016.

[50] M. R. DeLyser and W. G. Noid, "Extending pressure-matching to inhomogeneous systems via local-density potentials," *The Journal of Chemical Physics*, vol. 147, p. 134111, Oct. 2017.

[51] D. Rosenberger and N. F. A. van der Vegt, "Addressing the temperature transferability of structure based coarse graining models," *Physical Chemistry Chemical Physics*, vol. 20, no. 9, pp. 6617–6628, 2018.

[52] T. Sanyal and M. S. Shell, "Transferable Coarse-Grained Models of LiquidLiquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy," *The Journal of Physical Chemistry B*, Mar. 2018.

[53] M. Dallavalle and N. F. A. van der Vegt, "Evaluation of mapping schemes for systematic coarse graining of higher alkanes," *Physical Chemistry Chemical Physics*, vol. 19, no. 34, pp. 23034–23042, 2017.

[54] T. T. Foley, M. S. Shell, and W. G. Noid, "The impact of resolution upon entropy and information in coarse-grained models," *The Journal of Chemical Physics*, vol. 143, p. 243104, Dec. 2015.

[55] Y. Han, J. F. Dama, and G. A. Voth, "Mesoscopic coarse-grained representations of fluids rigorously derived from atomistic models," *The Journal of Chemical Physics*, vol. 149, p. 044104, July 2018.

[56] M. A. Webb, J.-Y. Delannoy, and J. J. de Pablo, "Graph-Based Approach to Systematic Molecular Coarse-Graining," *Journal of Chemical Theory and Computation*, Dec. 2018.

[57] M. S. Shell, "The relative entropy is fundamental to multiscale and inverse thermodynamic problems," *The Journal of Chemical Physics*, vol. 129, p. 144108, Oct. 2008.

[58] M. S. Shell, "COARSE-GRAINING WITH THE RELATIVE ENTROPY," in *Advances in Chemical Physics* (S. A. Rice and A. R. Dinner, eds.), pp. 395–441, Hoboken, NJ, USA: John Wiley & Sons, Inc., Sept. 2016.

[59] N. F. A. van der Vegt, C. Peter, and K. Kremer, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ch. 25, pp. 379–397. Chapman and Hall/CRC Press, Taylor Francis Group, Sept. 2008.

[60] A. Lyubartsev, A. Mirzoev, L. Chen, and A. Laaksonen, "Systematic coarse-graining of molecular models by the Newton inversion method," *Faraday Discuss.*, vol. 144, pp. 43–56, 2010.

[61] D. Rosenberger, M. Hanke, and N. F. van der Vegt, "Comparison of iterative inverse coarse-graining methods," *The European Physical Journal Special Topics*, vol. 225, pp. 1323–1345, Oct. 2016.

[62] S. Jain, S. Garde, and S. K. Kumar, "Do Inverse Monte Carlo Algorithms Yield Thermodynamically Consistent Interaction Potentials?," *Industrial & Engineering Chemistry Research*, vol. 45, pp. 5614–5618, Aug. 2006.

[63] N. J. H. Dunn, T. T. Foley, and W. G. Noid, "Van der Waals Perspective on Coarse-Graining: Progress toward Solving Representability and Transferability Problems," *Accounts of Chemical Research*, vol. 49, pp. 2832–2840, Dec. 2016.

[64] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé-Hoover chains: The canonical ensemble via continuous dynamics," *The Journal of Chemical Physics*, vol. 97, pp. 2635–2643, Aug. 1992.

[65] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Molecular Physics*, vol. 87, pp. 1117–1157, Apr. 1996.

[66] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, pp. 1701–1718, Dec. 2005.

[67] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," *Bioinformatics*,

vol. 29, pp. 845–854, Apr. 2013.

[68] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, "Versatile Object-Oriented Toolkit for Coarse-Graining Applications," *Journal of Chemical Theory and Computation*, vol. 5, pp. 3211–3223, Dec. 2009.

[69] S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle, and C. Junghans, "Relative Entropy and Optimization-Driven Coarse-Graining Methods in VOTCA," *PLOS ONE*, vol. 10, p. e0131754, July 2015.

[70] L. Zhang and J. I. Siepmann, "Pressure Dependence of the VaporLiquidLiquid Phase Behavior in Ternary Mixtures Consisting of n -Alkanes, n -Perfluoroalkanes, and Carbon Dioxide," *The Journal of Physical Chemistry B*, vol. 109, pp. 2911–2919, Feb. 2005.

[71] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, Oct. 1984.

[72] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, pp. 255–268, June 1984.

[73] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, pp. 7182–7190, Dec. 1981.

[74] G. Deichmann, M. Dallavalle, D. Rosenberger, and N. F. A. van der Vegt, "Phase Equilibria Modeling with Systematically Coarse-Grained Models-A Comparative Study on State Point Transferability," *The Journal of Physical Chemistry B*, vol. 123, pp. 504–515, Jan. 2019.

[75] `https://github.com/DRosen285/Pressure-matching`.

[76] N. J. H. Dunn, K. M. Lebold, M. R. DeLyser, J. F. Rudzinski, and W. Noid, "BOCS: Bottom-up Open-source Coarse-graining Software," *The Journal of Physical Chemistry B*, vol. 122, pp. 3363–3377, Apr. 2018.

[77] S. Plimpton, "Fast Parallel Algorithms for Short-Range Molecular Dynamics," *Journal of Computational Physics*, vol. 117, pp. 1–19, Mar. 1995.

[78] C.-C. Fu, P. M. Kulkarni, M. Scott Shell, and L. Gary Leal, "A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties," *The Journal of Chemical Physics*, vol. 137, p. 164106, Oct. 2012.

[79] R. A. Kullback, Solomon and Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86.

[80] W. F. van Gunsteren, X. Daura, and A. E. Mark, "Computation of Free Energy," *Helvetica Chimica Acta*, vol. 85, pp. 3113–3129, Oct. 2002.

[81] C. H. Bennett, "Efficient estimation of free energy differences from Monte Carlo data," *Journal of Computational Physics*, vol. 22, pp. 245–268, Oct. 1976.

[82] A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis, "Systematic determination of order parameters for chain dynamics using diffusion maps," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 13597–13602, Aug. 2010.

[83] A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, "Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach," *Chemical Physics Letters*, vol. 509, pp. 1–11, June 2011.

Transferability of local density assisted implicit solvation models for homogeneous fluid mixtures

David Rosenberger,[*,†] Tanmoy Sanyal,[‡] M. Scott Shell,[*,‡] and Nico F. A. van der Vegt[†]

†*Eduard Zintl Institut für Anorganische und Physikalische Chemie, Technische Universität Darmstadt, Darmstadt, Germany*

‡*Department of Chemical Engineering, University of California Santa Barbara, Santa Barbara, California, United States*

E-mail: rosenberger@cpc.tu-darmstadt.de; shell@engineering.ucsb.edu

Reproduced with permission from Journal of Chemical Theory and Computation, Article, ASAP DOI:10.1021/acs.jctc.8b01170.

Abstract

The application of bottom-up coarse grained (CG) models to study the equilibrium mixing behavior of liquids is rather challenging, since these models can be significantly influenced by the density or the concentration of the state chosen during parametrization. This dependency leads to low transferability in density/concentration space and has been one of the major limitations in bottom-up coarse graining. Recent approaches proposed to tackle this shortcoming range from the addition of thermodynamic constraints, to an extended ensemble parametrization, to the addition of supplementary terms to the system's Hamiltonian. To study fluid phase equilibria with bottom-up CG models, the application of local density (LD) potentials appears a promising approach, as shown in previous work by Sanyal and Shell [T. Sanyal, M. S. Shell, J. Phys.

Chem. B, 2018, 122, 5678]. Here, we want to further explore this method and test its ability to model a system which contains structural inhomogeneities only on the molecular scale, namely solutions of methanol and water. We find that a water-water LD potential improves the transferability of an implicit-methanol CG model towards high water concentration. Conversely, a methanol-methanol LD potential does not significantly improve the transferability of an implicit-water CG model towards high methanol concentration. These differences appear due to the presence of cooperative interactions in water at high concentrations that the LD potentials can capture. In addition, we compare two different approaches to derive our CG models, namely, relative entropy optimization and the Inverse Monte Carlo method, and formally demonstrate under which analytical and numerical assumptions these two methods yield equivalent results.

1 Introduction

Our understanding of the driving forces behind processes in soft condensed matter has greatly benefited from computer simulations, and from molecular dynamics (MD) simulations in particular. This technique allows an atomistic view into complex systems, but is limited by computational overhead to modeling length and time scales of tens of nanometers and microseconds, respectively. This hurdle can be overcome with coarse-grained (CG) particle models whose number of degrees of freedom (DOFs) and interactions to be evaluated are significantly smaller compared with fine-grained (FG) models. Apart from their computational advantages in computer simulations, CG models, and methods to derive them, provide important additional merits. Efforts to make models as simple as possible provide additional insight into emergent driving forces that may not be as easily obtained based on FG models alone. Systematic removal of DOFs involved in deriving a CG model based on its FG counterpart is the basis for bottom-up or systematic coarse graining.[1–3] To reduce DOFs, the high resolution, or FG, configuration space is projected onto a CG configuration

space lower in resolution. In order to evaluate the corresponding free energy surface, the so called multibody potential of mean force (PMF) must be calculated[2]

$$W(\mathbf{R}) = -k_B T \ln \int_V d\mathbf{r}\ \exp[-\beta U_{FG}(\mathbf{r})]\ \delta(\mathbf{R} - \mathbf{M}(\mathbf{r})) \tag{1}$$

where $W(\mathbf{R})$ is the multibody PMF and $\beta = 1/k_B T$ with k_B the Boltzmann constant and T the temperature. $\mathbf{M}$ is the projection, or mapping, operator, which relates a FG configuration $\mathbf{r}$ to a CG one $\mathbf{R}$. Due to the highly multibody nature of the integral, $W(\mathbf{R})$ is too complicated to compute exactly in practice, and only approximate solutions are possible. Solving an inverse problem is one way to obtain such an approximate solution to the multibody PMF. The goal is then to find a CG model that accurately reproduces one or multiple quantities of the FG model in the CG configuration space, by minimizing the difference between the FG and CG configuration space with respect to the quantity chosen. This inverse problem can be solved by means of either a variational principle[4–9] or by application of iterative Newton or quasi-Newton inversion techniques.[10–13]

Despite the methodological differences, all bottom-up coarse grained models suffer from the same two fundamental problems: accurate representability and state point transferability. The representability characterizes the ability of a CG model to simultaneously reproduce multiple properties of the FG system like structure, pressure and isothermal compressibility. The transferability instead describes the applicability of CG models at state points not included in the parametrization. Whereas accurate representability is guaranteed for the target property chosen to solve the inverse problem, it is not guaranteed for other properties, i.e. a match in structure does not automatically guarantee a match in thermodynamics or dynamics, or a match in forces does not guarantee structural agreement.[14–18] An equally outstanding challenge is the transferability of CG models.[19–22] The main reason why it is difficult to achieve both is that entropic contributions from particles "lost" upon coarse graining are missing. Therefore, the generated effective pair potentials cannot capture changes in

entropy necessary to describe certain thermodynamic properties and necessary to be transferable to different state points.[16,23,24] Several approaches have been proposed to tackle both representability and transferability. Among the most common are the addition of thermodynamic constraints to account for accurate pressure,[12] Kirkwood-Buff integrals[25] or the surface tension,[26,27] the application of an extended ensemble parametrization,[28,29] or the use of a extended Hamiltonian description for the energy of the system.[30–35]

Allen and Rutledge proposed the idea of local density (LD) dependent interactions to improve implicit solvent models.[20] Interactions based on the LD of CG sites have been used to improve the transferability[36] and the representability[37] of CG models. The local density of a CG site is simply a weighted local co-ordination number around that site, and can be written down generally for arbitrary combinations of central and neighboring types of sites. The LD potential can then be cast as a function of the local density and added as a corrective extension to the traditional pair-wise form of the Hamiltonian in CG models. Unlike pairwise Hamiltonians, a LD potential incorporates information about the inherently multibody environment around CG sites which contributes to enhanced model transferability. Further, it is a *mean-field* potential with computational complexity similar to that of pair potentials and thus does not sacrifice computational speed (further implementation details can be found in Ref [32]). LD potentials have been used to improve the sampling of conformation space in implicit solvent models of superhydrophobic polymers and were found to enhance the model's transferability to different polymer lengths.[32] Recently, they have also been shown to improve structural transferability in CG models of liquid mixtures such as benzene in water.[34]

In this work we want to test the ability of LD dependent CG models to quantitatively and qualitatively describe mixtures of water and methanol. These mixtures provide an interesting test case for the LD dependent potentials, since they show strong microheterogeneities at atomistic length scales as a function of methanol concentration, while remaining miscible at a macroscopic scale.[38–43]Laaksonen et al. showed that these microheterogeneities are caused

by a non-homogeneous distribution of the two components in the mixture.[44] This is expressed through structural patterns determined by the dominant component in the system. X-ray emission spectroscopy experiments applied by Guo et al. revealed that the inhomogeneous mixing between methanol and water can be explained by the formation of rings of methanol bridged by water molecules.[41] In agreement, Perera et al. found methanol molecules forming chain-like structures, caused by water bridging the hydroxyl groups of the methanol molecules.[45] Further, Pascal and Goddard confirmed the picture of incomplete mixing.[46] At low methanol concentrations, methanol molecules bury their hydrophobic groups away from water. With increasing methanol concentration this is no longer possible and free mixing is observed. At high methanol concentrations, the system is best described as water dissolved in methanol.[46]

Classical MD simulations of methanol-water mixtures have almost exclusively been based on all-atom force field models.[38,44–46] It is, however, interesting to ask if CG particle models can equally well be used to describe the structural properties of these systems. This question may provide insights to modeling large-scale phenomena driven by the interplay between hydrophobic and hydrophilic interactions in complex systems not amenable to all-atom models. In this study, we investigate the possibility to study this interplay with simple CG models. We do this by explicitly accounting for LD effects and examine if LD potentials can effectively describe the micro-heterogeneities observed in water/methanol mixture with a simple single site CG model for both liquids, water and methanol in an implicit solvent environment. We derive two different CG models: (I) CG methanol in implicit water and (II) CG water in implicit methanol. Further, we compare two different methods to generate bottom up CG models, namely Inverse Monte-Carlo (IMC)[11] and relative entropy optimization,[13] and we show analytically and numerically under which assumptions these two methods are equivalent.

The remainder of the article is structured as follows: first, the basic theoretical background on IMC and relative entropy optimization is given. Second, we prove analytically under

which assumptions IMC and the relative entropy method are equivalent. Next, we briefly discuss the extension of the relative entropy method to LD potentials, followed by the details of the numerical calculations performed. Adjacent, we present the main results of this study along with a detailed discussion, followed by the conclusion.

2 Methods

2.1 Inverse Monte Carlo

The Inverse Monte Carlo (IMC) Method, or Newton inversion method (introduced by Lyubartsev and Laaksonen) aims to derive a CG force field (FF) that reproduces the pairwise structure, i.e. the radial distribution function (RDF), of the underlying atomistic or fine grained (FG) system.[11] The CG FF is estimated initially as the two-body potential of mean force (PMF, $U^0(r_{ij})$) acting between two particles i and j along the distance r_{ij} obtained from the corresponding RDF of the FG system ($g^0(r_{ij})$):

$$U^0(r_{ij}) = -k_B T \ln g^0(r_{ij}) \tag{2}$$

In many cases, the PMF does not accurately resemble the effective pair potential in the CG configuration space due to the relevance of higher order correlations. Thus, the potential is updated a series of times n, solving a set of linear equations until the difference in the RDF is minimized. This leads to numerical pair potentials.

The set of linear equations is given by Eq.(3), where N_α is the number of particle pairs at a distance α either in the CG system or in the mapped reference system (N_α^0), $\boldsymbol{J}$ is a Jacobian matrix, ΔU is the potential update and γ is a particle pair distance $> \alpha$. The Jacobian matrix is defined in Eq.(4), where N_γ is the number of particle pairs separated by a distance

γ. Finally, the update of the PMF gets computed according to Eq.(5).

$$\langle N_\alpha \rangle^{CG} - N_\alpha^0 = \boldsymbol{J}_{\alpha,\gamma} \Delta U_\gamma \tag{3}$$

$$\boldsymbol{J} = \frac{\partial \langle N_\alpha \rangle^{CG}}{\partial U_\gamma} = -\beta(\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle) \tag{4}$$

$$U^n(r_{ij}) = U^{n-1}(r_{ij}) - \boldsymbol{J}^{-1}(g^{n-1}(r_{ij}) - g^0(r_{ij})) \tag{5}$$

The number of particle pairs N_α is related to the g(r) via[47]

$$\langle N_\alpha \rangle = \frac{N(N-1)}{2} \frac{4\pi r_\alpha^2 \Delta r}{V} g(r_\alpha) \tag{6}$$

where V is the volume of the system, N the number of particles and Δr is the discretization grid spacing. The Jacobian defined in Eq.(4) explicitly contains cross correlations between the number of particle pairs at different distances. This provides several advantages in converging towards the final effective pair potentials, and some disadvantages in terms of numerical stability of IMC, compared to the similar Iterative Boltzmann Inversion method[10,12] as discussed in the literature.[26,48,49]

2.2 Relative entropy optimization for pair potentials

Another way to determine a CG FF is through relative entropy optimization, as proposed by Shell.[13] The relative entropy is a quantity that measures the information loss upon reducing the FG system to the CG model. It is defined as

$$S_{\text{rel}} = \sum_i p_0(i) \ln \frac{p_0(i)}{p_{CG}(i)} + S_{\text{map}} \tag{7}$$

where p is the probability to observe a certain configuration i determined either by the FG FF (p_0) or by the CG FF (p_{CG}) and S_{map} is a mapping entropy which accounts for the degeneracy of atomistic states in the CG configuration space. In the canonical ensemble the

relative entropy can be expressed as

$$S_{\mathrm{rel}} = \beta\langle U_{CG}(\boldsymbol{\zeta}) - U_{FG}\rangle_{FG} - \beta(A_{CG}(\boldsymbol{\zeta}) - A_{FG}) + S_{\mathrm{map}} \tag{8}$$

where U is the potential energy, A is the Helmholtz free energy and $\boldsymbol{\zeta}$ is a vector that contains all parameters of the CG FF. The idea in relative entropy optimization is to find the optimal set of parameters $\boldsymbol{\zeta}$ that minimize the information loss between the FG and CG system. Similar to IMC, where the pair potential U is updated, in relative entropy optimization the FF parameters $\boldsymbol{\zeta}$ are updated until a minimum in the relative entropy is reached.[50] Using a Newton-Raphson approach, the parameter update scheme is:

$$\boldsymbol{\zeta}^{k} = \boldsymbol{\zeta}^{k-1} - \boldsymbol{H}^{-1}\nabla S_{\mathrm{rel}} \tag{9}$$

In Eq.(9), $\boldsymbol{H}$ is the Hessian matrix and ∇S_{rel} is the gradient of the relative entropy. Sequential iterations successively bring the parameters to a local S_{rel} minimum. This scheme is only applicable if the Hessian is positive definite, otherwise a steepest descent or conjugate gradient optimization scheme is applied.[50,51] It is interesting to note that if U consists of splines or tabulated potentials (as in IMC), then the relative entropy has a single global minimum and Eq.(9) always applies.[51]

2.3 Equivalence between Inverse Monte Carlo and relative entropy optimization

The two methods, IMC and relative entropy optimization (when using a Newton-Raphson update scheme), are equivalent for the derivation of effective pair potentials if the following

applies:

$$\nabla S_{\text{rel}} = (\langle N_\alpha \rangle^{CG} - N_\alpha^0)) \tag{10}$$

$$\boldsymbol{H}^{-1} = \boldsymbol{J}^{-1} \tag{11}$$

For simplicity we assume a system with only one component, but the equations can be easily extended to multi component systems as well.

As proposed by Lyubartsev and Laaksonen,[11] we start with a discretized Hamiltonian (U) to describe the potential energy of the system

$$U = \sum_\alpha N_\alpha(q)\zeta_\alpha \tag{12}$$

where ζ_α is the pair potential and N_α is the exact number of particle pairs at distance α, given by:

$$\langle N_\alpha \rangle = \frac{\int dq N_\alpha \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda)}{\int dq \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda)} \tag{13}$$

The Jacobian in Eq.(4) is then given by

$$\frac{\partial \langle N_\alpha \rangle}{\partial \zeta_\gamma} = \frac{\partial}{\partial \zeta_\gamma} \left(\int dq N_\alpha \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda) \cdot \frac{1}{\int dq \prod_\lambda \exp(-\beta N_\lambda(q)\zeta_\lambda)} \right) \tag{14}$$

which by application of the chain rule results in:

$$\boldsymbol{J} = \frac{\partial \langle N_\alpha \rangle}{\partial \zeta_\gamma} = -\beta(\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle) \tag{15}$$

By inserting Eq.(15) in Eq.(3), the potential update, ΔU_γ, in IMC is computed by solving:

$$\langle N_\alpha \rangle^{CG} - N_\alpha^0 = (-\beta(\langle N_\alpha N_\gamma \rangle - \langle N_\alpha \rangle \cdot \langle N_\gamma \rangle))\Delta U_\gamma \tag{16}$$

where we define the left hand side (l.h.s) as

$$\Delta N_\alpha^{IMC} \equiv \langle N_\alpha \rangle^{CG} - N_\alpha^0 \tag{17}$$

Next, we evaluate the relative entropy in Eq.(8). Its derivative can be written as:

$$\nabla S_{\text{rel}} = \beta \left\langle \frac{\partial U}{\partial \zeta} \right\rangle_{FG} - \beta \left\langle \frac{\partial U}{\partial \zeta} \right\rangle_{CG} \tag{18}$$

Substituting U defined by Eq.(12), we obtain:

$$\nabla S_{\text{rel}} = N_\alpha^0 - \langle N_\alpha \rangle^{CG} \equiv \Delta N_\alpha^{S_{\text{rel}}} \tag{19}$$

In comparison with Eq.(17), the result of Eq.(19) reveals the following relation:

$$-\Delta N_\alpha^{S_{\text{rel}}} = \Delta N_\alpha^{IMC} \tag{20}$$

The Hessian matrix in Eq.(9) is given by

$$\boldsymbol{H} = \left\langle \frac{\partial^2 U}{\partial \zeta_\alpha \partial \zeta_\gamma} \right\rangle_{FG} - \left\langle \frac{\partial^2 U}{\partial \zeta_\alpha \partial \zeta_\gamma} \right\rangle_{CG} + \beta \left\langle \frac{\partial U}{\partial \zeta_\alpha} \frac{\partial U}{\partial \zeta_\gamma} \right\rangle_{CG} - \beta \left\langle \frac{\partial U}{\partial \zeta_\alpha} \right\rangle_{CG} \left\langle \frac{\partial U}{\partial \zeta_\gamma} \right\rangle_{CG} \tag{21}$$

where the first two terms vanish due to the linearity of U in the parameters ζ.[51] The linearization that is exploited here results from using cubic splines for the effective pair potentials in the relative entropy method. The fact that the CG pair potential is linear in its parameters (spline knots) is crucial to the success of a simple scheme like Newton-Raphson descent in discovering a global minimum on the relative entropy surface which then theoretically guarantees robust representability at least for the pair correlations. Using the Hamiltonian

defined in Eq.(12), the remainder is:

$$\boldsymbol{H} = \beta\langle N_\alpha N_\gamma\rangle - \beta\langle N_\alpha\rangle\langle N_\gamma\rangle \tag{22}$$

which leads to a similar relation as in Eq.(20)

$$\boldsymbol{H}^{-1} = -\boldsymbol{J}^{-1} \tag{23}$$

Taking these results and applying the definition for the potential update (see Eq.(5) and Eq.(9)) we end with the following relation:

$$\boldsymbol{H}^{-1}(-\Delta N_\alpha^{S_{rel}}) = -\boldsymbol{J}^{-1}\Delta N_\alpha^{IMC} \tag{24}$$

$$\boldsymbol{H}^{-1}\Delta N_\alpha^{S_{rel}} = \boldsymbol{J}^{-1}\Delta N_\alpha^{IMC} \tag{25}$$

This proves that under the assumption of a discretized Hamiltonian (tabulated in IMC and represented with splines in the relative entropy method), quasi-Newton optimization strategies like Newton-Raphson leads to an exact equivalence of the pair potentials obtained through IMC and relative entropy. Moreover this shows that relative entropy optimization leads to a match in the RDF between the FG and CG system without directly using it as a target quantity in the optimization process. It is important to note that the agreement between IMC and relative entropy minimization is likely to also hold when very knot-dense spline potentials are employed in the latter (instead of discretized ones), although the necessary knot density and discretization to observe quantitative agreement may be high.

2.4 Relative entropy optimization for local density potentials

Recently Sanyal and Shell applied the relative entropy approach not only for pair potentials, but also for so called local density (LD) potentials.[32,34] LD potentials (U_{LD}) account for the effect of neighboring particles on the effective pair potential (u_{pair}). This additional

contribution changes the total CG potential energy to:

$$U_{CG} = \sum_{i<j} u_{\text{pair}} + U_{LD} \tag{26}$$

U_{LD} is a sum over an unspecified function of the local density around each particle i in the system,

$$U_{LD} = \sum_{i} f(\rho_i) \tag{27}$$

where $f(\rho_i)$ is practically represented using cubic B-splines. Here, the local density, ρ_i, is the total number of neighboring particles within a specified and smoothed cutoff (r_c)

$$\rho_i = \sum_{j \neq i} \phi(r_{ij}) \tag{28}$$

where the indicator function (ϕ) adopts a value of 1 below an inner cutoff r_0 but continuously and quickly decays to 0 at r_c. The shape of this function is chosen to be computationally convenient and does not require the calculation of the absolute distance between pairs of particles (or any square root operations).

$$\phi(r) = \begin{cases} 1 & r \leq r_0 \\ c_0 + c_2 r^2 + c_4 r^4 + c_6 r^6 & r \in (r_0, r_c) \\ 0 & r \geq r_c \end{cases} \tag{29}$$

The difference between r_0 and r_c will be called Δ and is of the order of $0.1 - 0.12$ nm. The exact form of the coefficients c as well as more details on the LD potentials can be found in the original work by Sanyal and Shell and the recent extension to binary mixtures.[32,34]

3 Simulation Details

FG and IMC CG simulations are performed with the Gromacs-5.1.2 MD engine.[52,53] IMC potentials are generated with the VOTCA coarse graining package (version 1.4).[47,54] The relative entropy optimization is achieved with an in house code. To account for local density potentials, CG simulations are executed with a modified version of the LAMMPS simulation package that includes a custom local density potential.[55]

3.1 Fine grained simulations

All systems studied contain 5000 molecules in total, with different methanol mole fractions, x_M, of interest, namely 0.1, 0.5 and 0.9 in SPC/E water.[56,57] For FG methanol a Kirkwood-Buff based force field for united atoms is used.[58] Newton's equations of motion are integrated based on a leap-frog algorithm with a timestep of 1 fs. All systems are equilibrated for 2 ns at a constant pressure of 1 bar and at a constant temperature of 300 K (NPT condition). For both the barostat and thermostat, the weak coupling method of Berendsen is applied with a coupling constant of $\tau_p = 1$ ps for the barostat and $\tau_T = 0.5$ ps for the thermostat.[59] The barostat compressibility is $4.5 \cdot 10^{-5}$ bar^{-1}. For the short-range van der Waals interactions a cut-off of 1.2 nm is applied with a long-range dispersion correction. Electrostatic interactions are treated with the particle-mesh-Ewald method[60] with a real space cut-off of 1.2 nm and a grid size of 0.12 nm. Bonded interactions in methanol are constrained with the LINCS algorithm.[61,62] Periodic boundary conditions are applied in the x, y, and z directions. The short equilibration is followed by a 10 ns run under NPT conditions replacing the Berendsen barostat and thermostat with the Parrinello-Rahman barostat[63] and the Nosé-Hoover thermostat[64] respectively. The coupling constants are set to $\tau_p = 1$ ps for the barostat and $\tau_T = 0.5$ ps thermostat. All other parameters are kept the same during the short equilibration run. The average volume of the 10 ns NPT simulation is then used

for the final production run under constant volume (NVT) conditions. Besides the volume constraint, all other parameters are the same as during the 10 ns NPT run.

3.2 Implicit solvent model

To generate the CG model for the united atom model of methanol in implicit SPC/E water[56,57] and for SPC/E water[56,57] in implicit methanol, a 3 to 1 mapping scheme is applied, where each molecule is mapped to its center of mass as illustrated in figure 1.

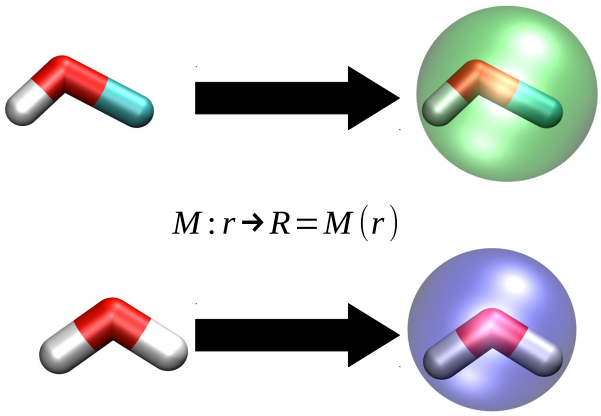


Figure 1: top: mapping scheme for CG methanol with the FG model on the left and the 1 bead representation (green) on the right; bottom: mapping scheme for CG water with the FG model on the left and the 1 bead representation (blue) on the right.

3.2.1 Inverse Monte Carlo optimization

On the basis of the FG RDFs between the centers of mass of water and water (WW) and methanol and methanol (MM) at $x_M = 0.5$, implicit solvent models for CG water and CG methanol are derived with the IMC method. As an initial guess for the iterative procedure, the PMF is taken (see Eq. (2)). To generate the CG configuration of the atomistic model,

water and methanol are mapped as illustrated in figure 1 and depending on the case – either water in implicit methanol or methanol in implicit water – the second component is implicitly present. This results in an uncharged single-site representation of the former 3 atom molecules. The effect of the lost component (either water in CG methanol or methanol in CG water) as well as the electrostatic interactions are integrated into the effective pair interaction generated via IMC. For the CG models of water in implicit methanol the RDFs are evaluated between 0.24 and 1.2 nm with a grid spacing of 0.01 nm. For CG methanol in implicit water the interval is changed to 0.3 and 1.5 nm with the same grid spacing. At each iteration a short MD simulation of 2 ns is performed with a timestep of 1 fs and with a leap-frog stochastic dynamics integrator. The simulations are performed at 300 K ((τ_T = 1.0 ps)) under NVT conditions using the average volume of the atomistic simulations at a mole fraction of 50% methanol ($x_M = 0.5$). No long-range dispersion correction for energy or pressure is applied for the numerical potentials during the CG simulation. The cut-off for the van der Waals interactions are the same as the maximum distance in the RDF evaluation. The iterative procedure is run until no change in the RDFs as well as in the potentials is further observed.

3.2.2 Relative entropy optimization

Relative entropy optimization is performed for the same CG models as described for IMC. The effects of the second component and the charges are again built into the effective interactions. The non-bonded pair potentials and the local density potentials are represented by cubic splines whose knot point values are determined during the optimization process. For both interactions 80 knot points are optimized. Models with only non-bonded pair interactions will be referred to as REO (Relative Entropy Optimization) models, whereas the case with pair splines and a local density potential will be called REO LD. The relative entropy at each iteration is calculated from short trial MD simulations launched with current estimates of the REO potentials. REO makes use of a reweighting scheme to decrease the number of trial MD runs thus reducing overall computation cost and statistical error.[50] During the

trial MD simulations the following steps are executed. First, the system is energy minimized with a conjugate gradient algorithm for 1000 steps, followed by 1 ns equilibration. Second, a 2 ns production run is performed under NVT conditions at 300 K. Newton's equations of motion are integrated according to the velocity-Verlet algorithm with a timestep of 1 fs. A Langevin thermostat is applied with a coupling constant of 0.1 ps The cut-off value for the REO potentials as well as the values for the outer cut-off (r_c in Eq. 29) and the difference between the outer and inner cut-off of the LD potentials (Δ) are listed in table 1.

Table 1: Cut-off values for the pair-potentials (r_c) and the outer cut-off values (r_c) for the LD potentials together with the difference between inner and outer cut-off for the LD potentials (Δ).

interaction	type	r_c(nm)	Δ (nm)
Methanol-Methanol (MM)	REO pair potentials	1.5	
	REO LD potentials	0.63	0.1
Water-Water (WW)	REO pair potentials	1.2	
	REO LD potentials	0.34	0.1

3.2.3 Coarse grained simulations

On the basis of the derived CG models, MD simulations are performed for 10 ns under NVT conditions at three different mole fractions of methanol $x_M = 0.1$, $x_M = 0.5$ and $x_M = 0.9$. The simulation parameters are the same as the ones used for the MD production phases in the iterative optimization approaches for IMC and relative entropy optimization.

4 Results and Discussion

4.1 Kirkwood-Buff Analysis of the fine grained system

In order to compare the derived CG models with their parent reference systems beyond structural accuracy, we also compute Kirkwood-Buff integrals (KBIs). KBIs relate local structure and thermodynamic properties like activity coefficients, solvation free energies or

the isothermal compressibility of stable mixtures.[65] For mixture components i and j they are defined as:[66]

$$G_{ij} = 4\pi \int_0^\infty [g_{ij}(s) - 1] s^2 ds \tag{30}$$

KBIs can be interpreted as the excess coordination number of particles j around a central particle i. This means that the larger the KBI value, the higher is the affinity between particles i and j. To evaluate Eq.(30) in computer simulations, a thermodynamic limit needs to be taken while the system should be open with respect to its components. Notwithstanding the latter requirement, KBIs can be calculated in computer simulations of closed (NVT or NPT) systems. To this end, the integral in Eq.(30) is usually truncated at an upper integration limit $r < L/2$ (with L the linear simulation box dimension) where $G_{ij}(r)$ (the running KBI or RKBI) is observed to oscillate around a mean, plateau value. In recent work, it has been shown that this mean value corresponds to the thermodynamic limiting value of Eq. (30). A more detailed discussion on the issue of finite size and ensemble effects is beyond the scope of this work and can be found elsewhere.[67–69]

By applying Eq.(30) one critical problem occurs that should not be ignored: RDFs do not strictly approach a limiting value of 1 in closed systems. This leads to a drift in the asymptotic behavior of the RKBIs. The drift is caused by depletion or accumulation of particles j around a particle i at local scales. This local depletion or accumulation is then compensated by a positive or negative excess of particles j at long distances, since the total number of particles j is constant. This leads to incorrect limiting behavior ($r \to \infty$) of $g_{ij}(r)$, which must be corrected. We here use the empirical correction introduced by Ganguly and van der Vegt[22]

$$g_{ij}^{corr}(r) = g_{ij}(r) \frac{N_j(1 - \frac{(4/3)\pi r^3}{V})}{N_j(1 - \frac{(4/3)\pi r^3}{V}) - \Delta N_{ij}(r) - \delta_{ij}} \tag{31}$$

where N_j is the number of particles j, V is the volume of the system, ΔN_{ij} is the excess number of particles j around a particle i within a sphere of radius r and δ_{ij} is the Kronecker

delta. KBIs obtained from $g_{ij}^{corr}(r)$ will be named Ganguly (GKBI) in the following.

Figure 2 shows the RKBIs of the FG simulations at $x_M = 0.5$. One clearly sees a drift in the asymptotic behavior of the water-water (WW) RKBI (solid black line) and how the empirical correction of Ganguly (solid red line) shifts the RKBI to larger values at longer distances. This is also observed for methanol-methanol (MM) (dotted lines). In case of methanol-water (MW, dashed lines), one observes that the RKBIs are already well converged without any correction. Here the empirical correction of Ganguly introduces a small shift in the tails of the KBIs towards lower values. Now if one compares the RKBI values averaged between 1.0 and 1.5 nm, the influence of the Ganguly correction is fairly small and the effect of the empirical correction becomes important only at larger distances.

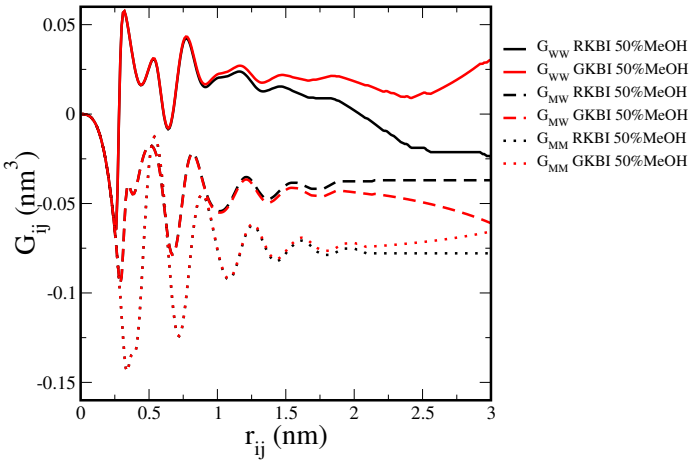


Figure 2: Influence of the Ganguly correction on the convergence of the Kirkwood-Buff Integrals in the AA system for Water-Water (solid lines), Methanol-Water (dashed lines) and Methanol-Methanol (dotted lines) at a mole fraction of x_M=0.5:uncorrected RKBI (black) and the RDF correction of Ganguly, GKBI, (red).

The average values for all KBIs obtained from FG simulations are presented in table 2. The averages are taken between 1.0 and 1.5 nm and the errors calculated by averaging over 5 independent simulations according to

$$err = \sqrt{\frac{\sum_i^N (G_{ij} - \langle G_{ij} \rangle)^2}{N(N-1)}} \tag{32}$$

where $N = 5$ is the number of simulations. As one sees, all KBI values are nearly the same no matter if the Ganguly correction is applied or not. The largest discrepancy between the RKBI and GKBI occurs for WW at $x_M = 0.9$. Further, one observes a minimum in G_{MM} at $x_M = 0.1$, which means that at low methanol concentrations the methanol molecules are mutually stronger depleted than at higher concentrations. This is in agreement with the work of Laaksonen et al.,[44] who showed that at low x_M the methanol molecules are separated by larger distances. Going from $x_M = 0.1$ to $x_M = 0.5$, G_{WW} and G_{MM} become larger (less negative). G_{WW} even changes sign and becomes positive. This indicates stronger water-water and methanol-methanol association in the $x_M = 0.5$ mixture. Upon further increasing the methanol content from $x_M = 0.5$ to $x_M = 0.9$, the values of G_{MW} in turn become larger while G_{WW} becomes smaller. Thus, at high mole fractions of methanol, the water molecules preferentially interact with methanol molecules. This observation agrees with the work of Pascal and Goddard.[46] The increase in G_{WW} upon raising the methanol concentration is also supported by other studies, which found that small water aggregates are formed the more methanol molecules are present in the system.[44–46] We note that the values of G_{MM} at $x_M = 0.1$, and of G_{WW} at $x_M = 0.9$, suffer from large uncertainties. This indicates that the KBIs are not fully converged and much longer sampling is needed.[22] For the purpose of this study we compare the FG model with the different CG models within the range of these uncertainties, being aware that a quantitative match might be difficult to achieve.

4.2 Effective potentials for the different implicit solvent models

The implicit solvent models are generated as described in the simulation details. The final CG potentials are presented in figure 3. Figure 3 a) shows the final effective pair potentials for the methanol-methanol (MM) interaction. As theoretically derived in the section on the equality between IMC and REO, REO (dashed green line) and IMC (solid red line) lead to very nearly the same set of pair potentials. The REO LD model shows a slightly smaller second maximum in the pair potential, which is illustrated through the dashed blue curve

Table 2: Kirkwood-Buff Integrals for methanol-methanol (MM), methanol-water(MW) and water-water (WW) at different methanol mole fractions (x_M) obtained by averaging RKBIs and GKBIs between 1.0 and 1.5 nm.

G_{ij}	x_M	RKBI (nm^3)	GKBI (nm^3)
G_{WW}	0.1	-0.026 ± 2.2·10^{-3}	-0.025 ± 2.5·10^{-3}
	0.5	0.021 ± 3.0·10^{-3}	0.025 ± 2.5 ·10^{-3}
	0.9	-0.02 ± 1.97·10^{-2}	0.01 ± 1.05 ·10^{-2}
G_{MW}	0.1	-0.039 ± 8.1 ·10^{-3}	-0.043 ± 8.7 ·10^{-3}
	0.5	-0.041 ± 1.3·10^{-3}	-0.043 ± 1.0 ·10^{-3}
	0.9	-0.016 ± 0.3 ·10^{-3}	-0.016 ± 0.4 ·10^{-3}
G_{MM}	0.1	-0.16 ± 3.58 ·10^{-2}	-0.14 ± 3.70 ·10^{-2}
	0.5	-0.078 ± 0.3·10^{-3}	-0.077 ± 0.4 ·10^{-3}
	0.9	-0.075 ± 0.2 ·10^{-3}	-0.075 ± 0.2 ·10^{-3}

in figure 3 a). This shift is caused by the additional MM LD potential, presented in figure 3 c). This potential is small in magnitude, but weakly attractive and lowers the potential barrier in the pair potential. Notably, the small magnitude of the LD potential reveals only a weak multibody nature of the MM interaction at this state point.

The CG potentials for water in implicit methanol are shown in figure 3 b). Similarly to CG methanol in implicit water, the effective pair potentials obtained from IMC (solid red line) and from the pair only relative entropy optimization (dashed green line) nearly overlap, with a small difference between the first maximum and the second minimum. Interestingly, the pair potential of the REO LD model (dashed blue line) lacks the inner potential well present in the other two models, but the absence of an attractive well is compensated by the additional WW LD potential (figure 3 d)). This points towards a strong coupling between the pair and LD potentials, which may mean that the attractive interactions are more naturally captured at a mean-field multibody level (LD potential), but this result may also be an outcome of an overlap in the function space of the pair and LD potentials whereby either can compensate for the another in a manner to which the CG optimization procedure is insensitive. Indeed similar compensation effects were reported by Sanyal and Shell[34] and more recently by Scherer and Andrienko for the coupling between two- and three-body

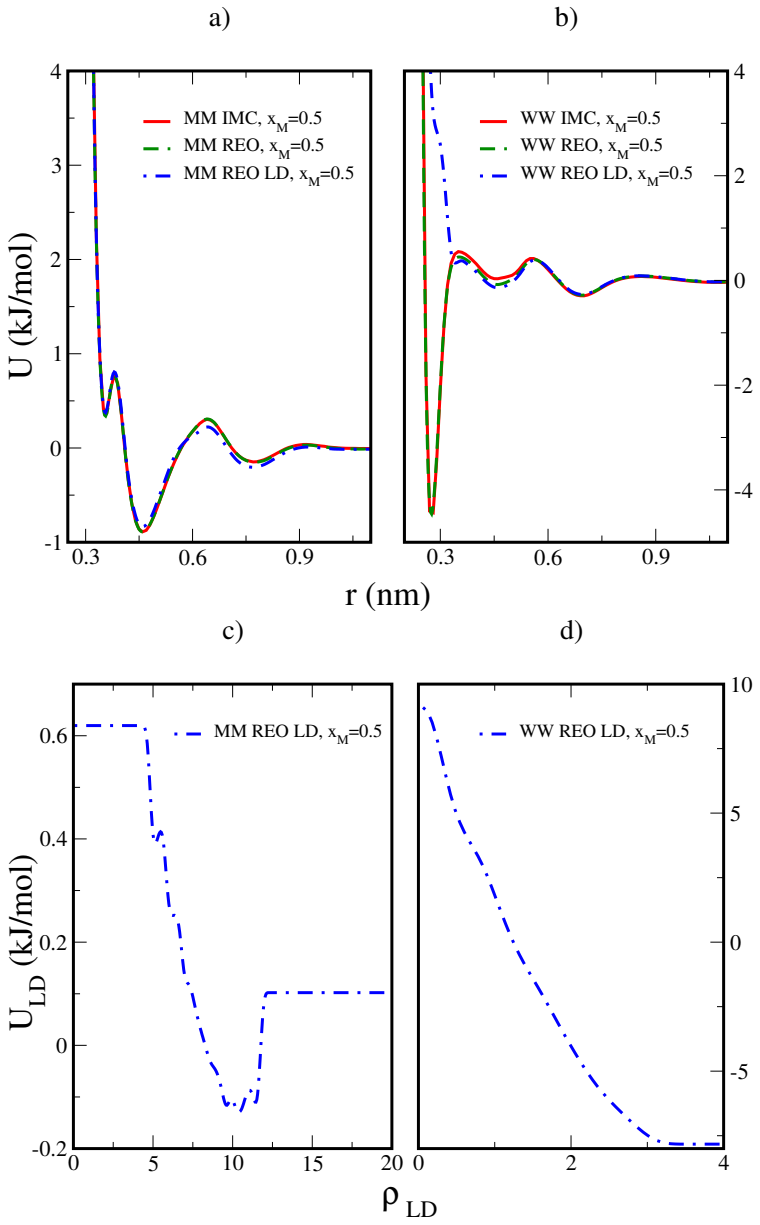


Figure 3: a) effective pair potentials for the methanol-methanol (MM) interaction at 50% methanol and b) for the water-water (WW) interaction obtained from IMC (solid red lines), REO (dashed green line) and REO LD optimization (dotted blue line); c) local density potentials for the MM interaction derived at 50% methanol and d) for the WW interaction.

interactions.[70] The WW LD potential is larger in magnitude compared to the MM one and saturates at a minimum of four neighboring water molecules, which no doubt corresponds to the preference for tetrahedral coordination in liquid water. The LD potential reaches a plateau from 4 neighbors onwards, which is by design of the relative entropy optimization. At some point the system no longer explores local densities beyond a maximal value (see for example figure 4 c)) where the largest is $\approx$ 15. So, beyond these values there is no information in the reference simulation that can be used to tune the potential. The relative entropy algorithm thus extrapolates a constant value thereafter. Noticeably, the WW IMC potentials needed 10 iterations more to converge than the MM ones (15 vs. 5 iterations).

4.3 Structural and Thermodynamic Representability

To evaluate the representability of the derived CG models, we compute the RDF between CG water in implicit methanol, and between CG methanol in implicit water, both at a mole fraction of x_M=0.5. In figure 4 a), the RDFs between methanol molecules for the different CG models in comparison with the center of mass RDF calculated from FG simulations are presented. As one sees, all CG models overlap with the FG RDF, showing two distinct maxima, one at $\approx$ 0.3 nm that stems from the methyl-hydroxyl interaction and one at $\approx$ 0.5 nm that stems from the methyl-methyl interaction. In figure 4 c), the LD distributions obtained from the REO and REO LD model are shown in comparison with the FG reference. As one sees, the REO model (dashed green line) already captures the FG LD distribution (solid black line) quite accurately and the additional LD potential improves the match only slightly (dashed blue line). This corresponds to the similarity in the underlying pair potentials and the small contribution of the LD potentials, and further indicates only a weak multibody nature of the MM interaction.

The water-water (WW) RDFs presented in figure 4 b), show a very sharp first peak that is present in all models. This points towards a high probability to find a water molecule next to another one, but due to the narrow width, only $\approx$ 4 water molecules are found in

the first solvation shell, in comparison to icosahedral coordination in simple liquids. This is in agreement with prior experimental and theoretical studies.[44,45,71] Despite the similarity in the qualitative appearance of the RDF, the REO LD model (dashed blue line) slightly underestimates this first peak, as shown in figure 4 b). The REO model (dashed green line) also slightly underestimates the first peak and the IMC model (solid red line) - by construction - shows the best agreement. This is surprising, since IMC and REO show only minor differences in the underlying pair potentials. One possible reason could be the different interpolation and extrapolation schemes applied by the two coarse graining software packages during the generation and/or evaluation of the final tabulated potentials. This assumption is supported by figure 5 a), where we compare the raw output of VOTCA, i.e. no interpolation between data points, and the two tabulated potentials used in the simulation. Here, potentials are interpolated. Moreover, the relative entropy approach utilizes piecewise cubic splines to represent interactions that are distinct from the potential interpolation in the IMC approach that uses the Akima interpolation scheme. Another reason for the differences between IMC and REO models may be the distinct convergence strategies, involving iteration in the former and minimization in the latter.

One sees that the repulsive part as well as the potential well in the tabulated IMC potential (solid black line) is slightly softer than for the REO tabulated potential (red line) and the raw output of VOTCA (green line). On the basis that structure is mainly determined by the short range part of a potential and given the narrow width of the first peak in the RDF, this small difference between raw output data and final tabulated potentials could cause the structural difference observed. For methanol-methanol, those small differences in the repulsive region are not observed as depicted in figure 5 b).

The methanol-methanol LD distributions seem insensitive to the parametrization method, consistent with the weak LD interactions that emerge in the REO LD model. The water-water LD distribution, on the other hand, reveals that the agreement between the FG model (solid black line) and the REO model (dashed green line) is improved by the LD potential

(dashed blue line), as shown in figure 4 d). Nevertheless, the improved agreement in the LD distributions comes with a loss of accuracy in the RDF, so it seems difficult to quantitatively match both at the same time with the current set of potentials. A reason for this may be the expanded parameter space for relative entropy minimization when LD potentials are included, in which spline discretization (i.e., knot density) begins to become important. This issue may also relate to the particular balancing of attractive interactions, which manifest in the LD strategy as a compensation between repulsive pair interactions and attractive LD ones, as discussed above. Indeed, in principle relative entropy minimization with pair and LD potentials should exactly reproduce both the pair and LD correlation functions, as discussed by Chaimovich and Shell,[72] such that observed differences must be due either to algorithm convergence, statistical fluctuations, or the manner by which the potential is approximated (e.g., splines of a chosen knot density).

To quantify the mismatch in the RDF, the root mean squared (RMS) error in the g(r) calculation between the FG and the CG system is computed as,

$$RMS\ \texttt{error} = \sqrt{\frac{\sum_{i=1}^{N}(g^{FG}(r) - g_i^{CG}(r))^2}{N}} \tag{33}$$

by performing the summation over a total of $N = 5$ simulations. The results are presented in figure 6. The error for the MM interaction, which is depicted in figure 6 a), is less than 0.01 for all the CG models, which further confirms the accurate structural quality of all CG methanol models in implicit water. However, for CG water the REO LD model (see figure 6 b)) shows a deviation of about 0.4 at the short distances, which corresponds to the underestimation of the first peak in the g(r) presented in figure 4 b). The IMC and REO models are very similar with respect to the RMS error and perform much better in terms of structural agreement compared to the REO LD model.

To further assess the quality of the derived implicit solvent CG models, the average GKBI values are computed for all models and compared to the FG system (see table 3). Note that these assessments are intimately connected to a model's representation of the pair

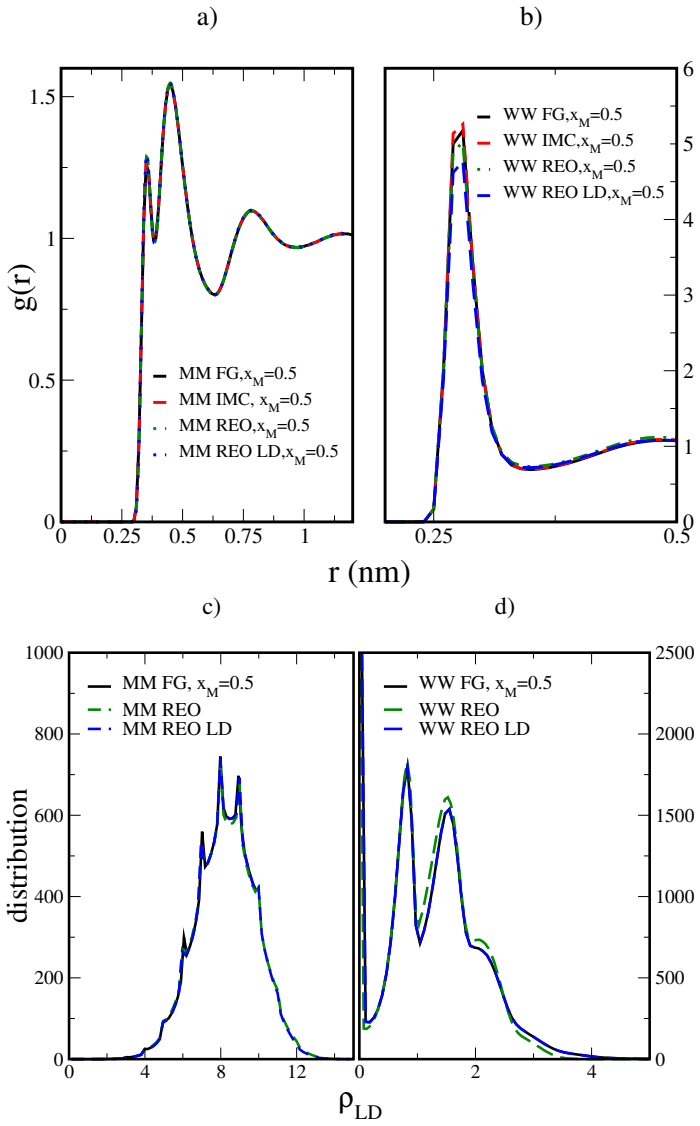


Figure 4: Representability analysis of the RDFs for the implicit solvent models at $x_M = 0.5$: a) Comparison of center of mass RDFs between methanol molecules (MM); b) Comparison of center of mass RDFs between water molecules (WW); Representability analysis of the LD distributions for the REO implicit solvent models at $x_M = 0.5$: c) Comparison of the LD distributions of methanol molecules (MM); d) Comparison of the LD distributions of water molecules (WW). The FG model is illustrated through the solid black line, the IMC model through the dashed red line, the REO model through the dotted green line and the REO LD model through the dotted blue line.

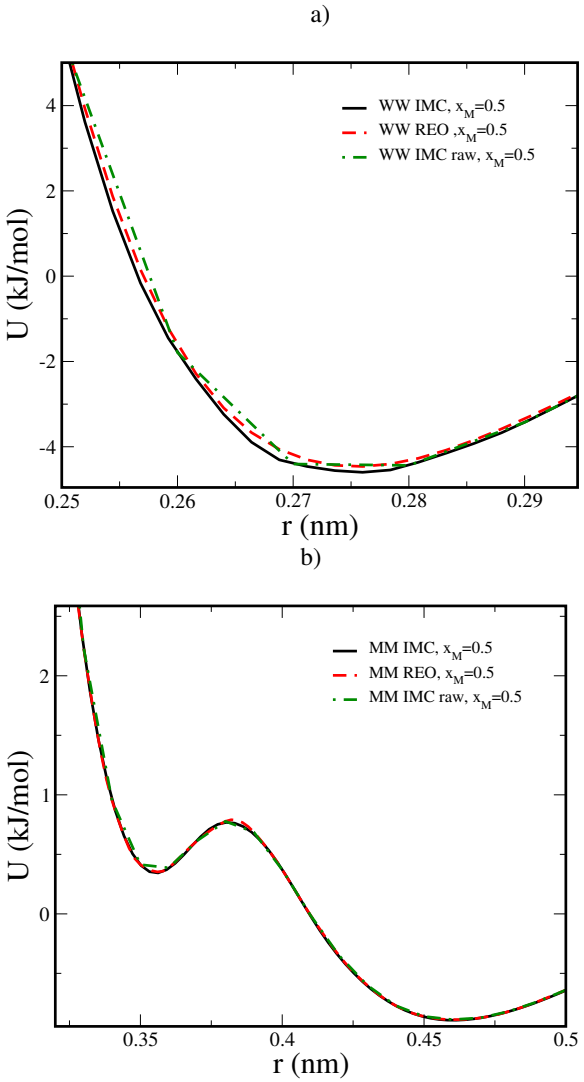


Figure 5: Comparison of final tabulated pair potentials between IMC (black line), REO (red line) and the raw output of the VOTCA simulation package (green line). a) for the water-water interaction; b) for the methanol-methanol interaction.

correlation functions, which all three methods should reproduce exactly in the limit of an infinite number of pair potential bins or spline points, according to the considerations in the section on the equality between IMC and REO. As such, differences in GKBI values between models necessarily represent differences in CG algorithm convergence properties and/or particulars of numerical representation of the potentials.

The GKBIs of all CG methanol models, G_{MM}, are in good agreement with the underlying FG one. This is in agreement with the structural overlap between all models. All CG water models show larger GKBI values compared to the FG reference. This general increase in the GKBIs for CG water in implicit methanol implies an effective stronger affinity between water molecules outside the first solvation shell. As a consequence, all CG water models are more compressible. The difference between the IMC and REO and REO LD model coincides with the differences observed in the RDFs. The REO and REO LD model show the same value within error bars. In general, inclusion of the LD potentials does not seem to impact the GKBI values, with respect to relative entropy pair-only models. This is likely due to the emphasis on long range correlations of the GKBI integrals and the fact that the LD potential modifies short range interactions.

Table 3: Ganguly corrected Running Kirkwood-Buff Integrals averaged between 1.0 and 1.5 nm for the different CG models

X_M	model	G_{WW} (nm^3)	G_{MM} (nm^3)
0.5	FG	$0.025 \pm 2.5 \cdot 10^{-3}$	$-0.077 \pm 0.4 \cdot 10^{-3}$
	IMC	$0.094 \pm 0.9 \cdot 10^{-3}$	$-0.078 \pm 0.2 \cdot 10^{-3}$
	REO	$0.063 \pm 2.2 \cdot 10^{-3}$	$-0.076 \pm 0.2 \cdot 10^{-3}$
	REO LD	$0.065 \pm 1.1 \cdot 10^{-3}$	$-0.074 \pm 0.1 \cdot 10^{-3}$

The similarity in the GKBIs between the REO and REO LD models might further point towards a dominant role of direct pairwise interactions in the system. To probe this idea, we compute the cluster size distribution that is an emergent property dependent on the collective interactions in the system. If multibody effects are significant, then one expects the cluster size distribution to be impacted by the ability of the CG model to reproduce them.[32]

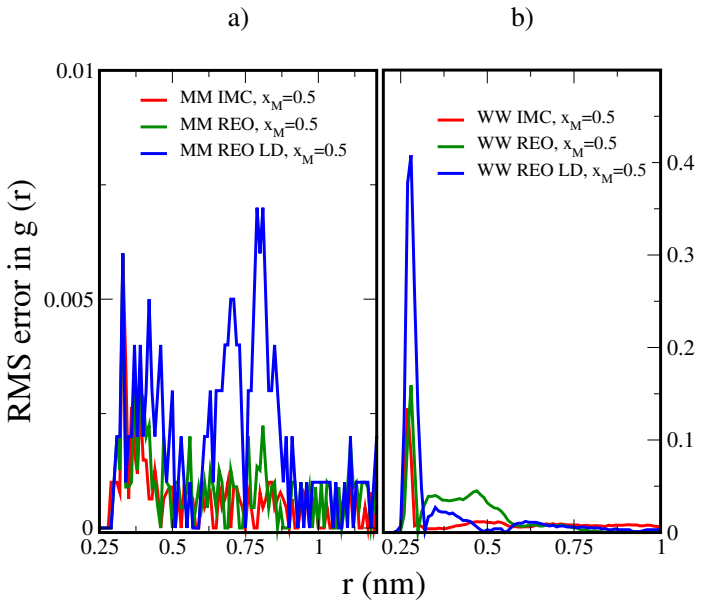


Figure 6: RMS error between FG and CG RDFs for the implicit solvent models at $x_M = 0.5$: a) RMS error in the methanol methanol (MM) RDF; b) RMS error in the water water (WW) RDF. The difference between the FG and the IMC model is illustrated through the solid red line, between the FG and the REO model through the solid green line and between the FG and the REO LD model through the solid blue line.

To evaluate the cluster size distributions, we chose the same cut-off values as for the LD potentials to determine whether or not particles are in contact and thus belong to the same cluster. In practice, the cluster size distribution can be very sensitive to the cutoff choice. As one sees in figure 7 a), the FG methanol molecules comprise a single, system-spanning cluster at $x_M = 0.5$. A corresponding snapshot of the FG system presented in figure 7 a) confirms this picture. Note that for visual clarity we do not show the water molecules, which are explicitly present in the simulation. The network-like structure is similar to the formation of rings and strings, which follows the observations made by Guo et al. and Perera et al..[41,45] Since all CG models are able to match the cluster size distribution of the underlying FG system, it suggests that multibody effects play a less significant role in forming this percolating structure.

Figure 7 b) shows the cluster size distribution of water molecules for the FG model and the implicit methanol models. Here, no large water clusters are formed and the water molecules largely populate isolated clusters as well as doublets and triplets, as indicated from the snapshot presented in the inset (methanol molecules are not depicted in the snapshot for visual clarity). The network-like methanol structure restricts the number of water-water contacts. At this concentration, water molecules are not significantly tetrahedrally coordinated, which supports the picture that the LD potential has a less attractive impact on the performance of the model compared to the REO or IMC model, since the number of neighboring molecules is most likely to be < 3.

Further, the formation of only small aggregates points towards a less dominant role of multibody effects, at least at the level of water-water interactions, and implies a weakened effect of the LD potentials. That observation most likely explains why the REO LD model is slightly less populated at low coordination numbers and cluster sizes. This observations coincides with the work of Laaksonen et al., who showed a loss in tetrahedral coordination of water molecules from $x_M = 0.5$ onwards.[44] The difference between the REO and REO LD model can be explained by an increased attraction between water molecules introduced by the neg-

ative tail of the LD potential if the number of neighboring waters is ≥ 3. This tail has a lower value than the minimum in the REO model and thus introduces stronger attraction.

4.4 Structural and Thermodynamic transferability

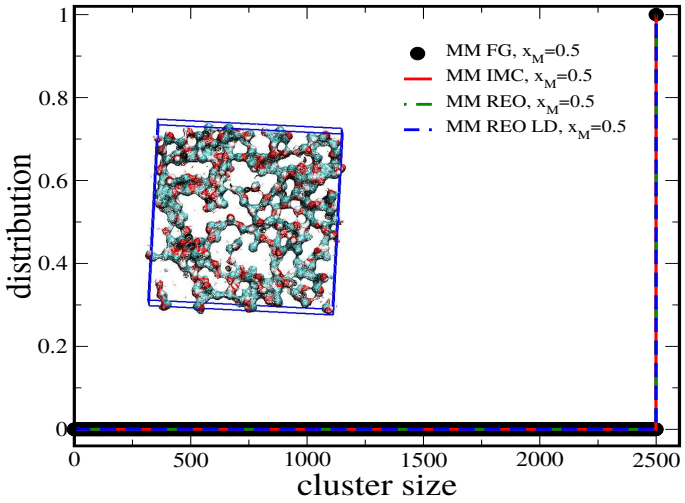
Despite the negligible contribution of LD potentials to the representability of the CG models at $x_M = 0.5$, we now test their effect on the transferability towards different methanol mole fractions. Because the IMC and REO optimized pair models are nearly identical, we focus on comparison of the IMC and the REO LD models to study the effect of LD potentials on the transferability of the derived CG models.

In the top panel of figure 8, the RDFs for the two CG models are shown at a mole fraction of $x_M = 0.1$ in comparison with the FG RDF at $x_M = 0.1$ and at $x_M = 0.5$. In figure 8 a), one can see that at $x_M = 0.1$ the CG models consistently show a decrease in the first peak of the MM RDF relative to the FG one at the same composition, and lower than the FG structure at $x_M = 0.5$. Further, the CG models underestimate the second maximum and overestimate the second minimum relative to the FG simulation at the same composition. Despite this quantitative mismatch, both CG models reproduce the main features (maxima, minima) of the FG structure well and show some aspects of transferability, when compared to the reference structure at $x_M = 0.5$.

For water, the REO LD model (solid light blue line) closely matches the structure of the FG system at high water content (solid black line), as depicted in figure 8 b). In contrast, the IMC water model fails to capture the g(r) structure and instead better seems to match the FG structure at $x_M = 0.5$ (solid orange line) that was the original parametrization condition. This may point to the role of multibody water-water interactions as relevant to the solution behavior at high water concentrations.

The bottom panel of figure 8 shows the RDFs at $x_M = 0.9$. In figure 8 c), the CG model for methanol in implicit water shows no significant difference between the IMC model (solid red line) and the REO LD model (dotted light blue line). Similar to the lower mole fraction case,

a)



b)

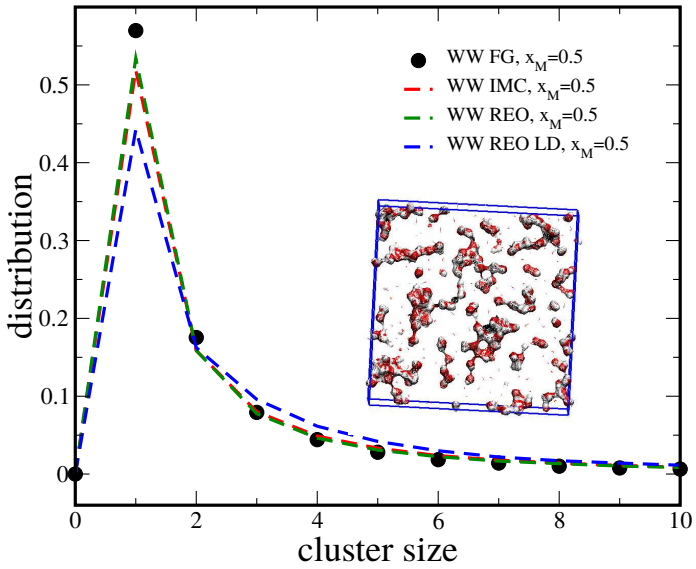


Figure 7: Representability analysis of the cluster size distributions for the implicit solvent models at $x_M = 0.5$: a) Comparison between the cluster size distribution of methanol molecules (MM). The inset depicts a snapshot of the FG trajectory without the water molecules for visual clarity; b) Comparison between the cluster size distribution of water molecules (WW). The FG model is illustrated through the black circles, the IMC model through the dashed red line, the REO model through the dotted green line and the REO LD model through the dashed blue line. The inset depicts a snapshot of the FG trajectory without the methanol molecules for visual clarity.

both of these models cannot reproduce exactly the FG structure (solid black line), but are able to capture the location of the first maximum in the RDF compared to the structure at at $x_M = 0.5$ (solid orange line). In figure 8 d), one sees that the IMC model (solid red line) leads to a decreased first peak in the RDF for CG water in implicit methanol compared to the FG model (solid black line). Further, the RDF obtained from the IMC model overlaps with the RDF of the FG system (dashed orange line) at the reference point ($x_M = 0.5$). These results suggest that IMC (and relative entropy minimization with pair potentials only) embed composition-specific interactions and correlations that propagate to other compositions. They also suggest that local density interactions can improve transferability when multibody interactions are present. On the other hand, the presence of the LD potential (solid light blue line) leads to a stronger water-water aggregation and a significantly enhanced first peak. It is likely in this case that multibody water interactions are over emphasized in dilute solutions where each water has few neighbors and the LD potential at low water-water coordination is inappropriate, having been constructed from a more water-rich reference.

In general, it seems that to model the structural transferability, LD dependent potentials play no significant role for methanol in implicit water. Contrarily, for water in implicit methanol, the presence of a LD potential improves the structural transferability with decreasing methanol concentration. This is in agreement with the observations of Sanyal and Shell,[34] who pointed out that water-water LD potentials lead to transferable CG models of aqueous mixtures only when water-water interactions have a major contribution to the multibody PMF. Further, this corresponds to the work of Laaksonen et al.,[44] who found out that the dominant species in the system is mainly responsible for the structural features of the mixture. Following this argument, the decreasing effect of LD contributions with increasing methanol concentration reveals the difficulty of using LD potentials to effectively capture the unusual mixing behavior of methanol and water upon increasing methanol mole fraction.

The negligible effect of multibody contributions is again illustrated by the methanol

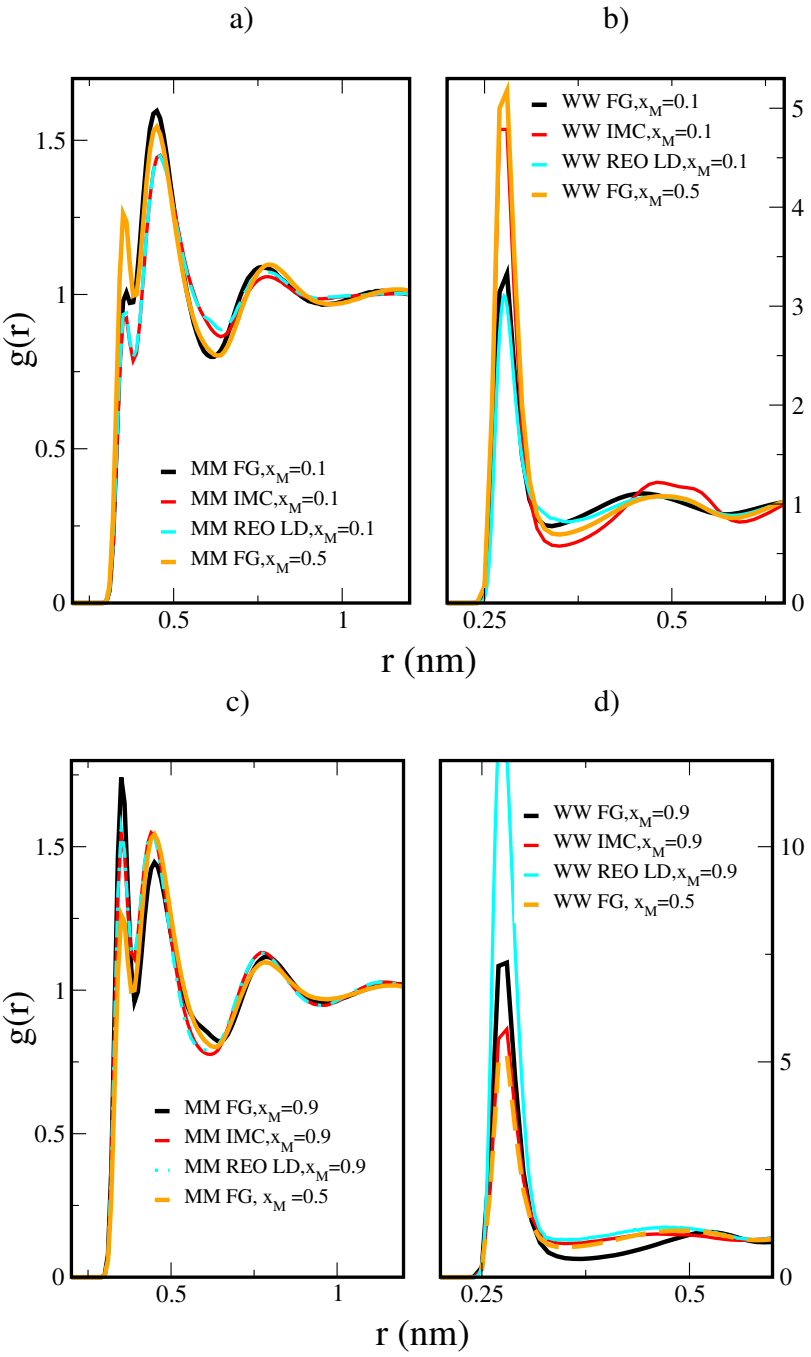
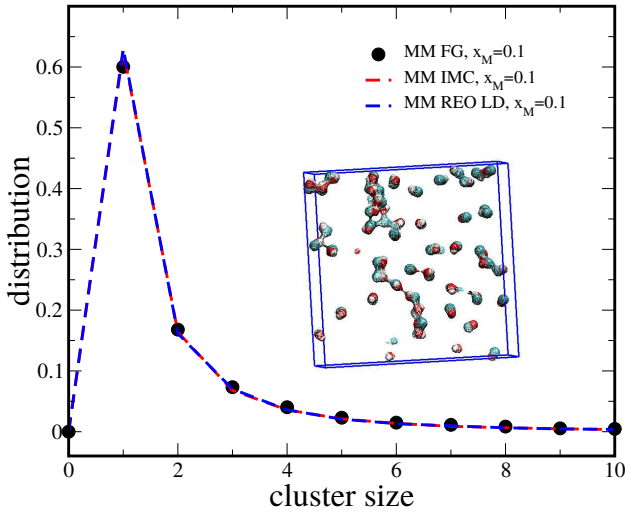


Figure 8: a) Comparison of the center of mass RDFs at $x_M = 0.1$ between methanol molecules (MM); b) Comparison of the center of mass RDFs at $x_M = 0.1$ between water molecules (WW); c) Comparison of the center of mass RDFs at $x_M = 0.9$ between methanol molecules (MM); d) Comparison of the center of mass RDFs at $x_M = 0.9$ between water molecules (WW). The FG model is illustrated through a solid black line, the IMC model is illustrated through the solid red line and the REO LD model is illustrated through the solid/dotted light blue line and the FG reference at $x_M = 0.5$ is illustrated through the solid orange line.

cluster size distribution for non-reference cases, as presented in figure 9. In figure 9 a), one sees that at low methanol concentrations only a few methanol molecules are grouped together, whereas at high concentrations (figure 9 b)) there exists only a single cluster composed of all methanol molecules in the system, which is qualitatively similar to x_M=0.5. These observations are further supported by the snapshots of the FG system depicted in the inset of the figures, where water molecules are not shown for visual clarity. There again a network-like structure is visible for methanol mole fractions > 0.5. Since all CG models perform in the same way, it further indicates that methanol-methanol interactions are not dominated by multibody effects in aqueous solutions.

We now turn to the cluster size distributions in the implicit-methanol systems. In figure 10 a), the distribution for waters at $x_M = 0.1$ shows two distinct cluster sizes for the FG system. On the one hand, a cluster emerges with very few water molecules, while on the other hand one appears with the number of water molecules in the system (i.e., a percolating or system spanning cluster). A snapshot of the FG system (shown in the inset) illustrates this, where the methanol molecules are not depicted for visual clarity. One sees a smaller cluster in the lower left corner and a large, system spanning cluster. It appears that small methanol sub-domains, present at low methanol concentration ($x_M = 0.1$, see figure 9 a)), prevent that all of the water aggregates into a single cluster. This produces the separation of peaks in the WW cluster distribution (see figure 10 a)) and is in agreement with the study of Perera et al..[45] The IMC model produces clusters with only few water molecules, that is more similar to the cluster distributions at $x_M = 0.5$. This is in agreement to what we observe for the g(r) as well, where the g(r) of the IMC model is more similar to the one of the FG model at $x_M = 0.5$ than at $x_M = 0.1$. Thus, this model is not transferable towards lower methanol concentrations. Interestingly, the REO LD model cluster distribution attains better agreement with the cluster size distribution, in particular capturing the two-state population at low and high cluster size. In this case, we adjusted the distance cut-off to 0.37 nm to effectively locate water neighbors in this heterogeneous environment. The cluster

a)



b)

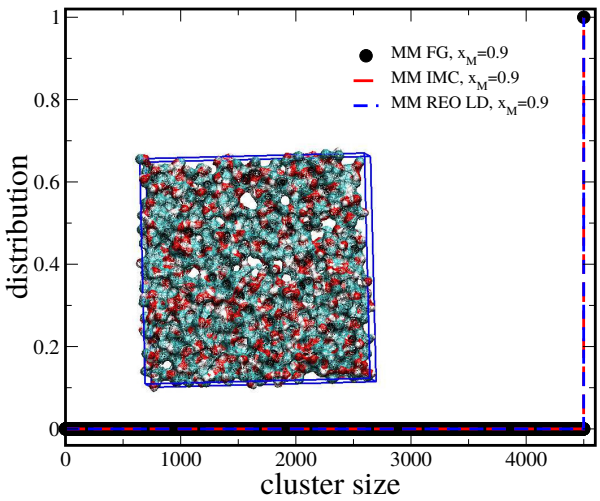


Figure 9: Transferability analysis of the cluster size distributions for the implicit solvent models: a): Comparison of cluster size distribution of methanol molecules (MM) at $x_M = 0.1$; b) Comparison at $x_M = 0.9$. The FG model is illustrated through black circles, the IMC model through the dashed red line and the REO LD model is illustrated through the dashed blue line. The insets depict a snapshot of the corresponding FG trajectories without the water molecules for visual clarity.

size distribution at $x_M = 0.1$ clearly shows that the LD potential improves the CG water model. Moreover, it confirms the picture that in water rich phases, multibody effects play a significant role and LD dependent potentials are effective strategies for capturing such interactions. At high methanol concentrations ($x_M = 0.9$), no large clusters of water are formed and the LD dependent potential does not improve the CG model compared to the IMC one. The more isolated behavior of water molecules is again depicted in a snapshot of the corresponding molecules taken from the FG system. This indicates a negligible role of water and multibody effects in methanol rich phases, which corresponds to the picture of Laaksonen and Perera.[44,45]

Another prediction of LD potentials is further indicated by the GKBIs presented in table 4. At low concentrations, both the IMC and REO LD model, show stronger affinity between methanol molecules than the FG model does. The LD potential even emphasizes this effect. The same trend can be observed at $x_M = 0.9$. Here, the REO LD model shows again the largest GKBI value, whereas the IMC model is closer to the FG system. This suggests further that LD effects do not play a significant role on the behavior of methanol in aqueous solutions.

Contrary to that, the REO LD model reproduces the FG WW-GKBI at low methanol concentrations ($x_M = 0.1$). At $x_M = 0.9$, the WW-GKBI is however overestimated by the REO LD model, as indicated by the RDF as well. Thus, the GKBI analysis follows the same trends observed in the RDFs. This brings us to the following conclusion: in water rich phases, where LD effects have a larger contribution to the water-water multibody PMF, LD potentials improve the transferability of the derived CG models. However, with increasing methanol concentrations, these contributions become less significant and the CG models do not further improve.

However, it is essential to note that the performance of all models in reproducing GKBIs is purely algorithmic in nature because the IMC, REO, and REO LD methods all formally should (in principle) locate CG models that reproduce exact g(r) forms upon which the

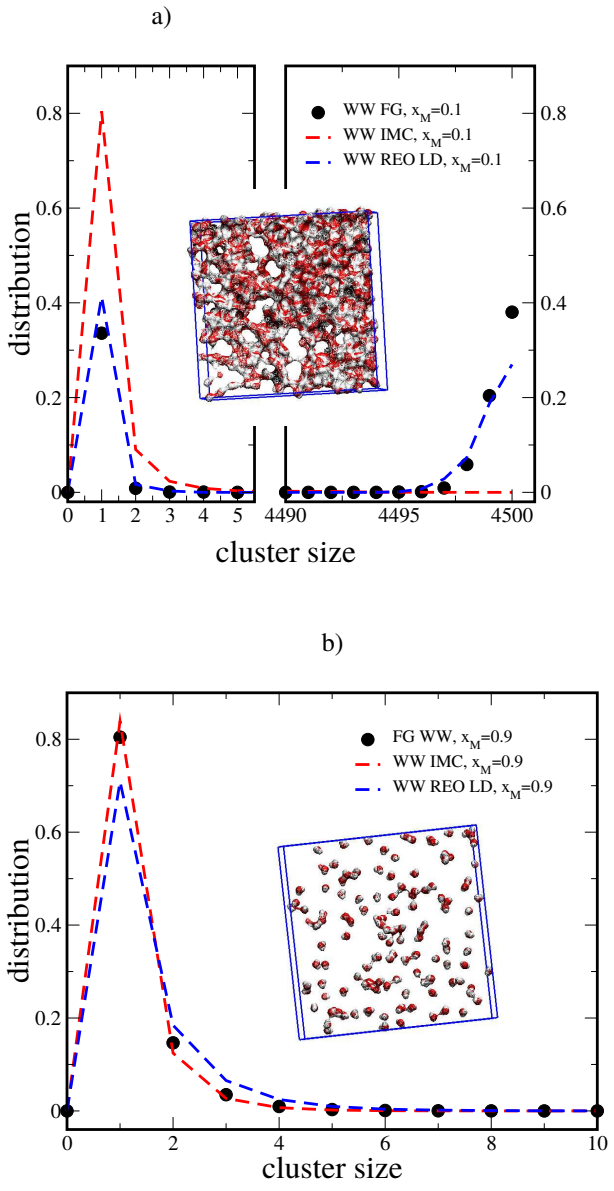


Figure 10: Transferability analysis of the cluster size distributions for the implicit solvent models: a) Comparison between the cluster size distribution of water molecules (WW) at x_M=0.1; b) Comparison between the cluster size distribution of water molecules (WW) at x_M=0.9. The FG model is illustrated through black circles, the IMC model through the dashed red line and the REO LD model is illustrated through the dashed blue line. The insets depict a snapshot of the corresponding FG trajectories without the methanol molecules for visual clarity.

GKBIs rely. Therefore, any difference between the properties in table 4 must be considered a result of numerical approximations (e.g., tabulated or splined potentials) or convergence (CG method), and the fact that the GKBI integrals are sensitive to subtle pair correlation effects at large distances.

Table 4: Average Ganguly corrected Running Kirkwood-Buff Integrals at different concentrations of methanol obtained either from FG simulations or from CG simulations with previously generated potentials at $x_M = 0.5$.

x_M	model	G_{WW} (nm^3)	G_{MM} (nm^3)
0.1	FG	-0.025 $\pm$ 2.5$\cdot 10^{-3}$	-0.14 $\pm$ 3.70 $\cdot 10^{-2}$
	IMC	0.090 $\pm$ 1.4$\cdot 10^{-3}$	-0.091 $\pm$ 1.2$\cdot 10^{-3}$
	REO LD	-0.026 $\pm$ 0.2$\cdot 10^{-3}$	-0.054 $\pm$ 0.2$\cdot 10^{-3}$
0.9	FG	0.01 $\pm$ 1.05$\cdot 10^{-2}$	-0.075 $\pm$ 0.3$\cdot 10^{-3}$
	IMC	0.084 $\pm$ 3.5$\cdot 10^{-3}$	-0.071 $\pm$ 0.2$\cdot 10^{-3}$
	REO LD	0.380 $\pm$ 0.6$\cdot 10^{-3}$	-0.066 $\pm$ 0.2$\cdot 10^{-3}$

5 Conclusion

In this work, we addressed the question of whether single site CG models for methanol and water mixtures, in an implicit solvent environment, can capture the mixing behavior of these two components. We could have also considered mapping methanol to a two-bead model to account for non-polar and polar sites. But, this would require up to 6 local density potentials for the combination of central and neighbor type. This would make the model (and its parameterization) significantly more complex, which we wanted to avoid. We derived implicit solvent CG models for various water/methanol mixtures by application of two different coarse graining methods, namely IMC and relative entropy optimization. We showed analytically and numerically, that under the assumptions of a discretized Hamiltonian and the application of a Newton-Raphson scheme, the methods are formally equal.

We further investigated if embedding of LD potentials improves the derived CG models in terms of concentration transferability. In agreement with previous work by Sanyal and

Shell[32,34] we found, if water-water LD contributions play a dominant role in the multibody PMF, LD potentials improve the transferability of water in implicit methanol towards lower methanol concentrations. Unlike the work in [32], here we find that LD potentials do not improve transferability in water-lean solutions. The LD included CG model for methanol in implicit water shows no significant improvement at either concentrations. When compared to a CG model with pair potentials only, the presence of a LD potential even emphasizes the affinity between methanol molecules as indicated by the GKBI. One difference in the present study from [34], which showed that LD models were consistently able to improve transferability across composition space in benzene-water mixtures, is that here a species is made implicit. It may be that LD potentials are more effective when all species are explicitly present but coarse-grained. The interpretation that multibody effects have a less significant effect on the structural arrangement of methanol molecules is further supported by the analysis of the cluster size distributions in the systems. Here, all CG models show similar size distributions for all concentrations independent of the presence of a LD dependent potential. The CG methanol models show more network-like structure due to large volume occupancy in the system, whereas the CG water stays mostly isolated, perhaps due to restraints posed by the methanol network. In agreement with our work, Scherer and Andrienko recently found a negligible effect of multibody contributions on methanol-methanol interactions by investigating the impact of three-body potentials on the pair structure of liquid methanol.[70] What seems to be more important than multibody contributions is the ability to form hydrogen bond networks in order to accurately describe water/methanol mixtures as discussed in the literature.[41,71] Explicit electrostatics or any directionality introduced to the model, as for example done on the basis of point multipole and Gay-Berne potentials,[73,74] seems to be necessary, to accomplish the effect of hydrogen bonding. Without performing extensive analysis on the FG system, we used bottom-up coarse graining techniques to identify the negligible contribution of LD dependent potentials on structural changes in alcohol/water mixtures. This does not mean that LD potentials cannot improve the transferability of

CG models as nicely demonstrated in the literature.[32–34] It rather shows that one should carefully consider the underlying physics of the system, specifically the extent of multibody effects, for a successful application of LD potentials. Further, this study highlights the aid of bottom-up coarse graining to identify important degrees of freedom, e.g. here the ability to explicitly form hydrogen bonds, to accurately describe a system in the CG configuration space.

Acknowledgement

NvdV and DR acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 233630050 - TRR 146 on "Multiscale Simulation Methods for Soft Matter Systems". TS and MSS acknowledge funding from the National Science Foundation through Grants Nos. CHEM-1300770 and CHEM-1800344.

References

(1) Peter, C.; Kremer, K. Multiscale simulation of soft matter systems from the atomistic to the coarse-grained level and back. *Soft Matter* **2009**, *5*, 4357.

(2) Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(3) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; van der Vegt, N. F. A. Systematic coarse-graining methods for soft matter simulations a review. *Soft Matter* **2013**, *9*, 2108–2119.

(4) Ercolessi, F.; Adams, J. B. Interatomic Potentials from First-Principles Calculations: The Force-Matching Method. *EPL* **1994**, *26*, 583–588.

(5) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

(6) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **2008**, *128*, 244114.

(7) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128*, 244115.

(8) Mullinax, J. W.; Noid, W. G. Generalized Yvon-Born-Green Theory for Molecular Systems. *Phys. Rev. Lett.* **2009**, *103*, 198104.

(9) Mullinax, J. W.; Noid, W. G. A Generalized Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials . *J. Phys. Chem. C* **2010**, *114*, 5661–5674.

(10) Schommers, W. Pair potentials in disordered many-particle systems: A study for liquid gallium. *Phys. Rev. A* **1983**, *28*, 3599–3605.

(11) Lyubartsev, A. P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730–3737.

(12) Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations: Mesoscale Potentials from Atomistic Simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636.

(13) Shell, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys.* **2008**, *129*, 144108.

(14) Louis, A. A. Beware of density dependent pair potentials. *J. Phys.: Condens. Matter* **2002**, *14*, 9187–9206.

(15) Johnson, M. E.; Head-Gordon, T.; Louis, A. A. Representability problems for coarse-grained water potentials. *J. Chem. Phys.* **2007**, *126*, 144509.

(16) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. On developing coarse-grained models for biomolecular simulation: a review. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423.

(17) Guenza, M. Thermodynamic consistency and other challenges in coarse-graining models. *Eur. Phys. J. ST* **2015**, *224*, 2177–2191.

(18) Wang, H.; Junghans, C.; Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur. Phys. J. E* **2009**, *28*, 221–229.

(19) Ghosh, J.; Faller, R. State point dependence of systematically coarse-grained potentials. *Mol. Simul.* **2007**, *33*, 759–767.

(20) Allen, E. C.; Rutledge, G. C. A novel algorithm for creating coarse-grained, density dependent implicit solvent models. *J. Chem. Phys.* **2008**, *128*, 154115.

(21) Brini, E.; van der Vegt, N. F. A. Chemically transferable coarse-grained potentials from conditional reversible work calculations. *J. Chem. Phys.* **2012**, *137*, 154113.

(22) Ganguly, P.; van der Vegt, N. F. A. Convergence of Sampling Kirkwood-Buff Integrals of Aqueous Solutions with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 1347–1355.

(23) Rudzinski, J. F.; Noid, W. G. Coarse-graining entropy, forces, and structures. *J. Chem. Phys.* **2011**, *135*, 214101.

(24) Wagner, J. W.; Dama, J. F.; Durumeric, A. E. P.; Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys.* **2016**, *145*, 044108.

(25) Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. A. Kirkwood-Buff Coarse-Grained Force Fields for Aqueous Solutions. *J. Chem. Theory Comput.* **2012**, *8*, 1802–1807.

(26) Murtola, T.; Falck, E.; Karttunen, M.; Vattulainen, I. Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints. *J. Chem. Phys.* **2007**, *126*, 075101.

(27) Moradzadeh, A.; Motevaselian, M. H.; Mashayak, S. Y.; Aluru, N. R. Coarse-Grained Force Field for Imidazolium-Based Ionic Liquids. *J. Chem. Theory Comput.* **2018**, *14*, 3252–3261.

(28) Mullinax, J. W.; Noid, W. G. Extended ensemble approach for deriving transferable coarse-grained potentials. *J. Chem. Phys.* **2009**, *131*, 104110.

(29) Moore, T. C.; Iacovella, C. R.; McCabe, C. Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion. *J. Chem. Phys.* **2014**, *140*, 224104.

(30) Das, A.; Andersen, H. C. The multiscale coarse-graining method. V. Isothermal-isobaric ensemble. *J. Chem. Phys.* **2010**, *132*, 164106.

(31) Dunn, N. J. H.; Noid, W. G. Bottom-up coarse-grained models that accurately describe the structure, pressure, and compressibility of molecular liquids. *J. Chem. Phys.* **2015**, *143*, 243148.

(32) Sanyal, T.; Shell, M. S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *J. Chem. Phys.* **2016**, *145*, 034109.

(33) DeLyser, M. R.; Noid, W. G. Extending pressure-matching to inhomogeneous systems via local-density potentials. *J. Chem. Phys.* **2017**, *147*, 134111.

(34) Sanyal, T.; Shell, M. S. Transferable Coarse-Grained Models of Liquid-Liquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy. *J. Phys. Chem. B* **2018**, *122*, 5678–5693.

(35) Rosenberger, D.; van der Vegt, N. F. A. Addressing the temperature transferability of structure based coarse graining models. *Phys. Chem. Chem. Phys.* **2018**, *20*, 6617–6628.

(36) Allen, E. C.; Rutledge, G. C. Evaluating the transferability of coarse-grained, density-dependent implicit solvent models to mixtures and chains. *J. Chem. Phys.* **2009**, *130*, 034904.

(37) Izvekov, S.; Chung, P. W.; Rice, B. M. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.* **2010**, *133*, 064109.

(38) Tanaka, H.; Gubbins, K. E. Structure and thermodynamic properties of water-methanol mixtures: Role of the water-water interaction. *J. Chem. Phys.* **1992**, *97*, 2626–2634.

(39) Soper, A. K.; Finney, J. L. Hydration of methanol in aqueous solution. *Phys. Rev. Lett.* **1993**, *71*, 4346–4349.

(40) Scatena, L. F. Water at Hydrophobic Surfaces: Weak Hydrogen Bonding and Strong Orientation Effects. *Science* **2001**, *292*, 908–912.

(41) Guo, J.-H.; Luo, Y.; Augustsson, A.; Kashtanov, S.; Rubensson, J.-E.; Shuh, D. K.; Ågren, H.; Nordgren, J. Molecular Structure of Alcohol-Water Mixtures. *Phys. Rev. Lett.* **2003**, *91*, 157401.

(42) Soper, A. K.; Dougan, L.; Crain, J.; Finney, J. L. Excess Entropy in Alcohol-Water Solutions: A Simple Clustering Explanation . *J. Phys. Chem. B* **2006**, *110*, 3472–3476.

(43) Laage, D.; Stirnemann, G.; Hynes, J. T. Why Water Reorientation Slows without

Iceberg Formation around Hydrophobic Solutes. *J. Phys. Chem. B* **2009**, *113*, 2428–2435.

(44) Laaksonen, A.; Kusalik, P. G.; Svishchev, I. M. Three-Dimensional Structure in Water-Methanol Mixtures. *J. Phys. Chem. A* **1997**, *101*, 5910–5918.

(45) Perera, A.; Zorani, L.; Sokoli, F.; Mazighi, R. A comparative Molecular Dynamics study of water-methanol and acetone-methanol mixtures. *J. Mol. Liq.* **2011**, *159*, 52–59.

(46) Pascal, T. A.; Goddard, W. A. Hydrophobic Segregation, Phase Transitions and the Anomalous Thermodynamics of Water/Methanol Mixtures. *J. Phys. Chem. B* **2012**, *116*, 13905–13912.

(47) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile Object-Oriented Toolkit for Coarse-Graining Applications. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.

(48) Lyubartsev, A.; Mirzoev, A.; Chen, L.; Laaksonen, A. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discuss.* **2010**, *144*, 43–56.

(49) Rosenberger, D.; Hanke, M.; van der Vegt, N. F. Comparison of iterative inverse coarse-graining methods. *Eur. Phys. J. ST* **2016**, *225*, 1323–1345.

(50) Carmichael, S. P.; Shell, M. S. A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly. *J. Phys. Chem. B* **2012**, *116*, 8383–8393.

(51) Shell, M. S. In *Advances in Chemical Physics*; Rice, S. A., Dinner, A. R., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016; pp 395–441.

(52) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(53) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(54) Mashayak, S. Y.; Jochum, M. N.; Koschke, K.; Aluru, N. R.; Rühle, V.; Junghans, C. Relative Entropy and Optimization-Driven Coarse-Graining Methods in VOTCA. *PLOS ONE* **2015**, *10*, e0131754.

(55) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(56) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(57) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(58) Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Methanol and Aqueous Methanol Solutions. *J. Phys. Chem. B* **2005**, *109*, 15080–15086.

(59) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(60) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(61) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(62) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

(63) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(64) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

(65) Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.* **1951**, *19*, 774–777.

(66) Ben-Naim, A. *Molecular theory of solutions*; Oxford University Phys. Rev. Ess: Oxford, 2006; OCLC: 252677203.

(67) Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. Kirkwood-Buff Integrals for Finite Volumes. *J. Phys. Chem. Letters* **2013**, *4*, 235–238.

(68) Dawass, N.; Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Simon, J. M.; Vlugt, T. J. H. Finite-size effects of Kirkwood-Buff integrals from molecular simulations. *Mol. Simul.* **2017**, 1–14.

(69) Milzetti, J.; Nayar, D.; van der Vegt, N. F. A. Convergence of Kirkwood-Buff Integrals of Ideal and Nonideal Aqueous Solutions Using Molecular Dynamics Simulations. *J. Phys. Chem. B* **2018**, *122*, 5515–5526.

(70) Scherer, C.; Andrienko, D. Understanding three-body contributions to coarse-grained force fields. *Phys. Chem. Chem. Phys.* **2018**, *20*, 22387–22394.

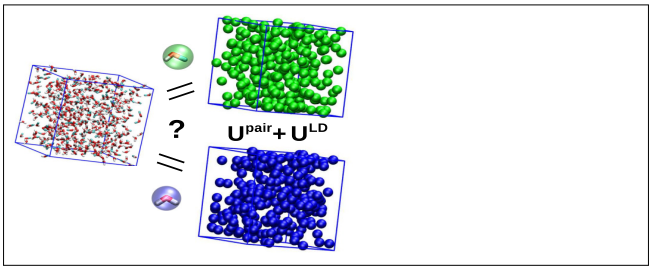
(71) Dixit, S.; Crain, J.; Poon, W. C. K.; Finney, J. L.; Soper, A. K. Molecular segregation observed in a concentrated alcohol-water solution. *Nature* **2002**, *416*, 829–832.

(72) Chaimovich, A.; Shell, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **2011**, *134*, 094112.

(73) Golubkov, P. A.; Ren, P. Generalized coarse-grained model based on point multipole and Gay-Berne potentials. *J. Chem. Phys.* **2006**, *125*, 064103.

(74) Golubkov, P. A.; Wu, J. C.; Ren, P. A transferable coarse-grained model for hydrogen-bonding liquids. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2050.

Graphical TOC Entry



PCCP

PAPER

Cite this: *Phys. Chem. Chem. Phys.*, 2018, **20**, 6617

Addressing the temperature transferability of structure based coarse graining models

David Rosenberger * and Nico F. A. van der Vegt

Received 8th December 2017,
Accepted 12th February 2018

DOI: 10.1039/c7cp08246k

rsc.li/pccp

Systematically derived coarse grained (CG) models for molecular liquids do not inherently guarantee transferability to a state point different from its reference, especially when derived on the basis of structure based CG methods like Inverse Monte Carlo (IMC). Several efforts made in the past years to improve the transferability of these models focused on including thermodynamic constraints or on the application of multistate parametrization. Das and Andersen (DA) [Das *et al.*, *J. Chem. Phys.*, 2010, **132**, 164106.] proposed a different Ansatz. They derived a correction term added to the system's Hamiltonian to reproduce the virial pressure and the volume fluctuations of the reference system in the CG resolution which does not require further adjustment of the effective pair potential. Herein, we discuss the possibility to achieve temperature transferability with IMC models for selected alkanes following the optimization of the DA approach as proposed by Dunn and Noid (DN) [Dunn *et al.*, *J. Chem. Phys.*, 2015, **143**, 243148.]. The work focuses on a novel approach to determine the DN correction term for different state points by linear interpolation.

1 Introduction

Systematic coarse graining of molecular systems involves a projection of a high dimensional, fine-grained (FG), configuration space onto a lower dimensional coarse-grained (CG) configuration space. This projection requires the selection of degrees of freedom (DOF), which are considered irrelevant for the physical problem of interest and are integrated out by the coarse graining procedure.[1–3] CG models facilitate linking chemistry and properties of complex systems on long time and large length scales, which are typically inaccessible in FG molecular dynamics (MD) simulations. Development of CG models by means of systematic coarse graining furthermore allows identifying relevant DOF that determine the representability of the CG model.

A possible way to systematically generate CG models is by application of so called structure based methods, whose goal it is to generate effective potentials between particles which reproduce the structure of the FG system in the CG representation. These structure-based methods are the iterative Boltzmann inversion (IBI),[4,5] the inverse Monte-Carlo (IMC) method[6] and the generalized Yvon–Born–Green (g-YBG)[7,8] version of the multiscale-coarse graining (MS-CG) approach.[9–12] IBI and IMC can be seen as Newton inversion techniques. Both methods generate effective pair potentials (EPP) by an iterative update of the potential of mean force (PMF) acting between two particles until a match in the radial distribution function (RDF) in the CG and mapped FG system is achieved. In IBI the PMF update is calculated by an approximate Jacobian,[13] whereas IMC uses an exact Jacobian[14] to calculate the potential update. The difference in the update term has consequences on the robustness and the convergence of the algorithms and thus on the final potentials. As a consequence IBI and IMC can provide different EPP, although both EPP reproduce the structure of the FG system.[15]

In contrast to IBI and IMC, the g-YBG approach generates a structure based CG force field (FF) by application of a variational principle.[16] This approach does not require an iterative update and is based only on structural correlation functions following the Yvon–Born–Green hierarchy for simple liquids.[17] Thus the g-YBG method accounts for 3-body contributions in the derivation of an effective 2-body potential. In contrast, IBI[4,5] and IMC[6] derive 2-body potentials based on 2-body correlations only. Nevertheless, Russ *et al.* have shown that IMC-type methods can be applied to derive 3-body interactions as well.[18] Despite it's higher complexity the g-YBG FF does not naturally reproduce any distribution function of the FG model, because of differences in the cross-correlations between the mapped FG and CG ensemble. Due to its lower resolution the CG FF might show less coupling between the different DOF than the mapped ensemble does.[16] Similar to IBI and IMC the g-YBG FF can be iteratively refined to improve the representability of certain correlation functions.[19] However good structural representability of a FG system does not automatically guarantee good representability of other thermodynamic and dynamic quantities like pressure, Kirkwood–Buff integrals (KBIs), thermal expansion or diffusion coefficients.[20,21] For example,

Technische Universität Darmstadt, Eduard-Zintl-Institut für Anorganische und Physikalische Chemie, Alarich-Weiss-Straße 10, 64287, Darmstadt, Germany. E-mail: rosenberger@cpc.tu-darmstadt.de

structure based CG models overestimate the virial pressure of the system, since the attractive part of the EPP is not exactly described by these models.[22] This is because the structure is mainly determined by the repulsive part of the EPP. The lack of cohesive energy between the particles causes an expansion of the system if a barostat is applied to keep the pressure of the CG system equal to that of the FG system. Moreover transferability remains one of the biggest challenges in bottom-up coarse graining, since there is no obvious relation between representability and transferability to state points different than the one chosen during parametrization.[23–27]

In order to improve both, thermodynamic representability and the transferability of structure based CG models, there are several approaches which have been applied and discussed in the literature. In the context of this work we want to focus ourselves on IBI/IMC-type methods. IBI models have been improved by the addition of linear ramp potentials, which shift the EPP in order to reproduce a second quantity in addition to the RDF. This second quantity can be either a pressure, which improves the representability of the virial pressure of the FG system,[5] or KBIs which enhance the transferability over a range of concentrations.[28] Moore *et al.* showed by application of a multistate parametrization further, that IBI potentials are transferable to states of different densities.[29] Here the EPP are determined by applying IBI over multiple states such that a state and time averaged potential is obtained.[29] Temperature transferability of IBI models was achieved by Farah *et al.* due to application of a linear interpolation scheme between EPP derived at two different temperatures.[30] All these different methods involve either a change of an already parametrized potential in terms of a linear ramp or development of a set of potentials for at least two state points.

In the context of IMC the application of a ramp correction might be counteractive, since it can cause convergence problems of the algorithms.[31] Nevertheless Murtola *et al.* showed that it is possible to directly include thermodynamic constraints when the potential update is calculated.[31] In their study, these authors have investigated the effect of using the surface tension as a constraint to derive EPP for a phospholipid/cholesterol bilayer system. Including the surface tension as a constraint for the potential update improved the convergence of the method and showed transferability towards larger system sizes, but not towards different system compositions.

To improve the representability of IMC-type methods, Almarza *et al.* used the potential energy as a constraint of the potential update.[32] In their work, the effective pair potential got updated not only based on the mismatch in the RDF, but also based on the mismatch in the potential energy per particle. On the basis of this difference, the effective pair potential got uniformly shifted, which improved the structural representability of the derived model for liquid aluminium.

In this work we want to address the question, if it is possible to achieve temperature transferability with IMC models for liquid alkanes with neither a thermodynamic constraint nor by application of a multistate parametrization. To do so, we follow the Ansatz of Das and Andersen (DA)[33] in order to get a representable and temperature transferable CG model obtained by IMC. This approach has been developed in the framework of the MS-CG method, *i.e.* a force based coarse graining method. DA proposed to add a volume dependent potential (U_V) to the system's Hamiltonian which accounts for the correct virial pressure and the correct volume fluctuations of the systems.[33] The idea of adding a volume dependent term to the Hamiltonian of the system goes back to the late 1970s, when Ashcroft and Strout discussed this idea in the context of liquid metals.[34] Later, volume dependent terms have also been added to the Hamiltonian of soft matter systems like charged colloids,[35] or mixtures of free colloids and non-adsorbing polymers.[36] A brief overview on the topic can be found for example in a review article by Likos.[37] In the framework of DA, U_V is determined in a two-step process. First, a purely configurational CG FF is generated, *i.e.* the position dependent interactions between the particles, by any bottom-up coarse graining method of choice, here IMC. Second, U_V is determined by a variational principle such that difference between the virial pressure of the FG system and the virial pressure produced by the configurational CG FF is minimized. We will refer to this method as the Das–Andersen pressure matching (DA-PM) approach from now on. As shown by DA, the additional U_V term does not only allow simulations with an external barostat, but further gives transferability towards different system sizes. The approach of DA was followed by Dunn and Noid (DN), who could show, by coarse graining liquid alkanes, that a self-consistent correction of U_V has to be applied in order to get a representable[38] and a concentration transferable model.[39] This self-consistent approach, which requires an optimization of U_V at each state point, will be called the Dunn–Noid pressure matching (DN-PM) method. Recently DeLyser and Noid extended this method.[40] They combined the DN-PM method with local density dependent potentials to model inhomogeneous systems. In their approach the interaction potential between particles is described by the sum of EPP and a term dependent on the local environment of a particle following the approach of Sanyal and Shell.[41] This local density dependent potential is then coupled with the DN-PM approach.

To the best of our knowledge both PM-approaches, DA and DN, have never been combined with an iterative inverse structure based coarse graining method like IMC nor has the temperature transferability of CG models been investigated with PM methods. That is why we discuss a novel combination of IMC with the DA-PM and DN-PM approach and investigate the temperature transferability for two model molecular liquids, namely hexane (HEX) and perfluorohexane (PFH). We show that it is possible to achieve temperature transferability with decent accuracy without any further changes made to the EPP. Further we show that it is sufficient to apply the self-consistent DN-PM approach only at two state points and to use a linear regression (LR) scheme to inter/extrapolate to state points not included in the parametrization. This novel approach, called Dunn–Noid linear regression (DN-LR), allows a fast and efficient parametrization to obtain a temperature transferable CG model, since the effective interactions have to be determined only once. This is different from the approach of Farah *et al.*, who also used a

linear interpolation scheme to achieve temperature transferability for a CG model of liquid HEX.[30]

The article is structured in the following way: first, a concise explanation of the theoretical background of the applied methods is given. Second, the simulation details are presented, followed by the main results and the corresponding discussion. In the end we conclude with a short summary and a short outlook on remaining challenges in the field.

2 Theoretical background

2.1 Derivation of a structure based coarse-grained force field *via* inverse Monte Carlo

The derivation of a CG FF with structure based bottom-up methods is an inverse problem. The true solution to this problem is an exact mapping of the FG FF onto the CG configuration space, where the corresponding potential is the so called many-body PMF.[3] This many-body PMF samples each variable in CG ensemble with the mapped probability distributions obtained from the underlying FG system. In reality the many-body PMF is too complex to get computed, thus only an approximate solution of the inverse problem is possible.[14] To generate such an approximate solution, the many-body PMF is split into bonded and non-bonded contributions. The bonded potential energy can be further separated into three different contributions[42]

$$U(r,\theta,\phi) = U_{\mathrm{r}}(r) + U_{\theta}(\theta) + U_{\phi}(\phi) \tag{1}$$

where $U_{\mathrm{r}}(r)$ is the potential energy which acts along the bond length r, $U_{\theta}(\theta)$ is the angular (θ) potential and U_{ϕ} is the torsional (ϕ) potential. To generate each of the bonded potentials in eqn (1) one can use the Boltzmann inversion (BI) scheme. BI calculates the PMF (U^0) of a degree of freedom q based on the mapped probability distribution ($P^0(q)$), where q can be any of the variables describing a bonded interaction.[42]

$$U^0(q) = -k_{\mathrm{B}}T \ln P^0(q) + C_{\mathrm{q}} \tag{2}$$

In eqn (2) k_{B} is the Boltzmann constant, T is the temperature and C_{q} is a constant.

The PMF obtained by BI should lead to a quantitative reproduction of the mapped FG probability distribution for a specific variable q in the CG ensemble. This is usually achieved for the bonded interactions, since they are less affected by higher order correlations present in the system. The non-bonded interactions can be derived based on BI of the RDF between two particles, but only in combination with an iterative refinement of the PMF until a match in the RDF between the CG and FG ensemble is achieved. This iterative procedure minimizes the indirect contributions of higher order correlations to accurately reproduce the pair structure of the FG ensemble. The IMC method of Lyubartsev and Laaksonen is an exact Newton method to calculate this refinement.[6] In IMC the PMF gets updated by a term ΔU_{γ}, which is obtained by iteratively solving a set of linear equations[43]

$$\langle N_{\alpha}\rangle^{\mathrm{CG}} - N_{\alpha}^{0} = \mathbf{A}_{\alpha,\gamma}\Delta U_{\gamma} \tag{3}$$

until the structural difference between the CG and FG system has reached a minimum (left hand side of eqn (3)). Here N_{α} is the number of particle pairs at a distance r_{α}, which is related to the RDF, $g(r_{\alpha})$, by[43]

$$\langle N_{\alpha}\rangle = \frac{N(N-1)}{2}\frac{4\pi r_{\alpha}^2 \Delta r}{V} g(r_{\alpha}) \tag{4}$$

where V is the volume of the system, N the number of particles and Δr is the shell thickness. The Jacobian matrix $\mathbf{A}$ in eqn (3) expresses the change of N_{α} with respect to small changes of the potential at a distance γ.[6]

$$\mathbf{A} = \frac{\partial\langle N_{\alpha}\rangle^{\mathrm{CG}}}{\partial U_{\gamma}} = -\beta\left(\langle N_{\alpha}N_{\gamma}\rangle - \langle N_{\alpha}\rangle\langle N_{\gamma}\rangle\right) \tag{5}$$

In eqn (5) β is $1/k_{\mathrm{B}}T$ and k_{B} is the Boltzmann constant and T is temperature. Since the Jacobian contains exact first order derivatives and cross-correlations, it has several consequences regarding the convergence and robustness of the method compared with the similar IBI method. A detailed discussion and comparison of these two methods is beyond the scope of this work and has been carried out in a previous study of the authors.[15]

2.2 Pressure matching

Despite structural agreement, IMC potentials overestimate the virial pressure of the system, since the EPP does not accurately describe the cohesive energy between the particles.[15] This leads to an expansion of the system, if simulations are performed in the isothermal and isobaric (*NPT*) ensemble at the pressure of the FG system. However, to investigate the thermal transferability of a CG FF, MD simulations have to be performed under *NPT* conditions, since the volume has to adjust to reproduce the correct bulk density at a chosen temperature and pressure. To avoid the problem of expansion, we follow the method of DA in order to get a good representable and temperature transferable structure based CG model.[33] In this article we briefly summarize the main idea of DA and present the key steps of the method. For a more profound discussion the interested reader is directed to the work of DA and DN.[33,38,44]

The DA-PM method introduces a purely volume dependent potential term $U_{\mathrm{V}}(V)$ to the system's Hamiltonian,[33]

$$H(\mathbf{R}^N, \mathbf{p}^N, V) = \sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{R}^N) + U_V(V) \tag{6}$$

where $\mathbf{R}^N$ is the position vector of all particles N, $\mathbf{p}_i$ is the momentum of a particle i, and $U(\mathbf{R}^N)$ is the configurational part of the CG FF. This part depends only on the positions of the particles and can be derived by BI (for the bonded part) and IMC (for the non-bonded part) as described under Section 2.1. The additional $U_{\mathrm{V}}(V)$ term contains information on the average virial pressure and the volume fluctuations of the CG system, it does not depend on the positions of the particles and is defined as:[33]

$$U_{\mathrm{V}}(V) = \frac{NV}{\langle V\rangle}\psi_1 + N\left(\frac{V}{\langle V\rangle} - 1\right)^2\psi_2 \tag{7}$$

where N is the particle number of the CG system, $\langle V\rangle$ the average volume of the FG system, V the volume of the FG system at time t and ψ_1 and ψ_2 are unknowns which have to be determined such that the virial and the volume fluctuations of the FG system are accurately reproduced by the Hamiltonian defined in eqn (6). The corresponding force is calculated as[33]

$$F_V = -\frac{\partial U_V}{\partial V} = -\frac{N}{\langle V\rangle}\psi_1 - 2N\frac{V-\langle V\rangle}{\langle V\rangle^2}\psi_2 \quad (8)$$

where the ψ_1-term accounts for the average correction of virial pressure (ΔP) and the ψ_2-term accounts for average correction of the inverse compressibility[38]

$$\Delta P = -\frac{N}{\langle V\rangle}\psi_1 \quad (9)$$

$$\Delta\kappa_T^{-1} = \frac{2N}{\langle V\rangle}\psi_2 \quad (10)$$

To determine the unknown coefficients ψ_1 and ψ_2 a variational principle is solved.[38]

$$\chi^2(U_V|U_R) = \langle|P_{FG}(\mathbf{r},\mathbf{p},V) - (P^0_{CG}(\mathbf{M}(\mathbf{r}),\mathbf{M}_\mathbf{p}(\mathbf{p})) + F_V(V))|^2\rangle \quad (11)$$

In eqn (11) $\mathbf{r}$ are the position coordinates of the particles in the FG system, $\mathbf{M}(\mathbf{r})$ are the mapped position coordinates in the CG resolution and $\mathbf{p}$ and $\mathbf{M}_\mathbf{p}(\mathbf{p})$ are the corresponding momenta. The solution of eqn (11) determines the corresponding values for ψ_1 and ψ_2, which minimize the difference between the FG virial pressure ($P_{FG}(\mathbf{r},\mathbf{p},V)$) and the virial produced by the configurational part of the CG FF applied on the mapped FG positions and momenta ($P^0_{CG}(\mathbf{M}(\mathbf{r}),\mathbf{M}_\mathbf{p}(\mathbf{p}))$). Note that the average in eqn (11) is performed over the mapped FG ensemble. The DA-PM approach is exact, if the CG model exhibits the same volume fluctuations as the FG model, since its parametrization is based on that (see eqn (11)). In case the configurational part of the CG FF leads to different volume fluctuations compared to the mapped FG ensemble, the DA-PM approach needs an iterative refinement as it is shown by DN.[38] The self-consistent DN-PM approach minimizes the difference in the pressure equation of state (EOS) between the FG and CG system given by[38]

$$\partial P(V) = P_{FG}(V) - P^0_{CG}(V) \quad (12)$$

where all variables have the same meaning as in eqn (11). The deviation defined in eqn (12) is minimized by determining ψ_1 and ψ_2 for both systems, the FG and the CG system, independently. On the basis of the difference in the pressure EOS, the coefficients determined by the DA-PM method get updated until the CG pressure EOS equals the FG one.

3 Simulation details

3.1 Fine grained simulations

FG simulations were performed with the Gromacs-5.1.2 MD package.[45,46] Both molecules, HEX and PFH, have been modeled with the TraPPE-united atom (ua) FF.[47] Each of the systems contained 500 molecules of either pure HEX or PFH. The molecules were packed into cubic boxes with periodic boundary conditions in x, y, and z direction. The box sizes were chosen to represent the bulk densities at each temperature of interest. For HEX the range of temperatures investigated was between 240 K and 320 K with a step size of 20 K and for PFH between 200 K and 340 K with steps of 20 K (200 K to 280 K) or 10 K (280 K to 320 K) in between. To integrate Newton's equations of motion a leap-frog algorithm with a timestep of 1 fs was applied. All systems were equilibrated for 5 ns at a constant pressure of 1 bar and at a constant temperature (*NPT* condition). Pressure and temperature were kept constant with the Berendsen barostat and thermostat with a coupling constant of $\tau_p = 1$ ps for the barostat and $\tau_T = 0.5$ ps for the thermostat respectively.[48] The barostat compressibility was 4.5×10^{-5} bar^{-1}. For the short-range van der Waals interactions a cut-off of 1.4 nm was applied with a long-range dispersion correction. For the bonded interactions no constraints were enforced. After the short equilibration, a 100 ns production run was performed. During the production run the Berendsen barostat and thermostat were replaced with the Parrinello–Rahman barostat[49] and the Nosé–Hoover thermostat[50] respectively. The coupling constants were set to $\tau_p = 5$ ps for the barostat and $\tau_T = 0.5$ ps thermostat. All other parameter were kept the same as during the short equilibration.

3.2 Generation of the coarse grained force field

The CG models were derived with the aid of the VOTCA package (version 1.4)[43,51] in combination with Gromacs-5.1.2.[45,46] The CG model of HEX is based on a 2 to 1 mapping scheme, which maps two connected HEX united atoms onto their center of mass (com) as shown on the left hand side in Fig. 1. The two terminal beads (colored green on the left hand side in Fig. 1) are referred to as beads of type A and the central bead (colored red on the left hand side in Fig. 1) is referred to as bead of type B. For PFH a 6 to 1 mapping scheme was applied based on the com of the whole molecule. The mapping scheme is depicted on the right hand side in Fig. 1. We call the single bead C (colored red on the right hand side in Fig. 1).

3.2.1 Potential generation. Based on the three-bead mapping scheme, the bonded interactions for HEX were derived *via* BI according to eqn (2). Therefore we took the probability distributions for the bond length of the A–B bond and for the A–B–A angle at 300 K. The non-bonded interactions were generated by application of the IMC method. In order to ensure faster convergence, we applied 30 steps of IBI to generate a better initial guess than the simple PMF. To further improve the robustness of the method, regions of poor sampling were removed from the target

Fig. 1 Left hand side: 3-bead mapping scheme for HEX: in green beads of type A, in red, bead of type B; right hand side: 1-bead mapping scheme for PFH: in red bead of type C.

distributions. The final grid for the evaluation of the RDFs was between 0.32 nm and 1.5 nm for the A–A interactions, between 0.33 nm and 1.4 nm for the B–B interactions and between 0.33 nm and 1.35 nm for the A–B interactions with a grid spacing of 0.01 nm. We performed 70 IMC iterations with a short 2 ns simulation (500 ps during the IBI steps) at each iteration. Newton's equations of motion were solved with a leap-frog stochastic dynamics integrator with a time steps of 1 fs. No long-range dispersion correction was applied during the iterations. The cut-off for the non-bonded CG interactions was 1.5 nm. All simulations were performed at 300 K under *NVT* conditions with an inverse friction constant of $\tau_T = 1.0$ ps.

The CG model of PFH was generated based on the RDF between the com of 2 PFH molecules, which was evaluated between 0.38 and 1.5 nm with a grid spacing of 0.01 nm. We again performed short simulations of 2 ns with a timestep of 1 fs at each iteration, where all the simulation parameter were the same as for the HEX simulations. For PFH we applied 10 steps of preliminary IBI, which were followed by the final 10 steps of IMC.

3.3 Pressure matching

The unknown coefficients in eqn (8), ψ_1 and ψ_2, were determined from the mapped FG trajectories of both systems, HEX and PFH, at each temperature studied. We applied the – rerun option of Gromacs-5.1.2[45,46] to compute the virial pressure. During the rerun, the FG FF was replaced with the previously determined IMC (non-bonded part)/BI (bonded part) FF. In case of HEX all other parameters were kept consistent with the FG simulation presented in Section 3.1. Whereas for PFH the long range dispersion correction term was turned off during the rerun. Based on the virial pressure obtained from the rerun the variational principle, as defined in eqn (11), was solved using the numpy library for python. The obtained coefficients were iteratively updated according to DN-PM method to minimize the error defined in eqn (12). Therefore CG simulations were performed with a modified version of the LAMMPS simulation package[52] which contains the DA correction term $U_V(V)/F_V(V)$. A code for the PM and a modified version of LAMMPS are available within the recently released Bottom-up Open-source Coarse-graining Software (BOCS).[53] The calculations were performed under *NPT* conditions with the Martyna–Tobias–Tuckerman–Klein barostat with a Nosé–Hoover chain length of three for the thermostat.[54,55] The coupling constant for the thermostat was $\tau_T = 0.2$ ps and for the barostat respectively $\tau_P = 1$ ps. Simulations were run for 10 ns with a time step of 1 fs. After each simulation a new set of ψ_1 and ψ_2 was determined until the FG virial and volume fluctuations were satisfactorily reproduced.

4 Results and discussion

4.1 Evaluation of the fine grained models

We performed MD simulations according to the set up presented in Section 3.1 to validate the quality of our reference models. The bulk densities obtained from the simulations are compared with experimental data[56] at different temperatures as shown in the top panel of Fig. 2. As we see, both FG models show accurate qualitative and quantitative agreement with the experimental densities in the range of temperatures we investigated. Hence, the TraPPE-ua FF for HEX and PFH is a suitable reference model to build a temperature transferable bottom-up CG model for each of the molecules. For further comparison with our CG models, we calculated the RDF between the com of HEX and PFH molecules respectively. In the central panel of Fig. 2 we show that the RDF between the com of HEX molecules (black line) contains a rather broad first peak with two distinct maximum values at 240 K, which vanish when the temperature gets increased (300 K red line and 320 K green line). The com RDF at the two higher temperatures, 300 K and 320 K, are very similar. In the bottom panel of Fig. 2 the RDF between the com of PFH molecules is presented. The com RDF of PFH rather shows a small shoulder than two distinct peaks at the

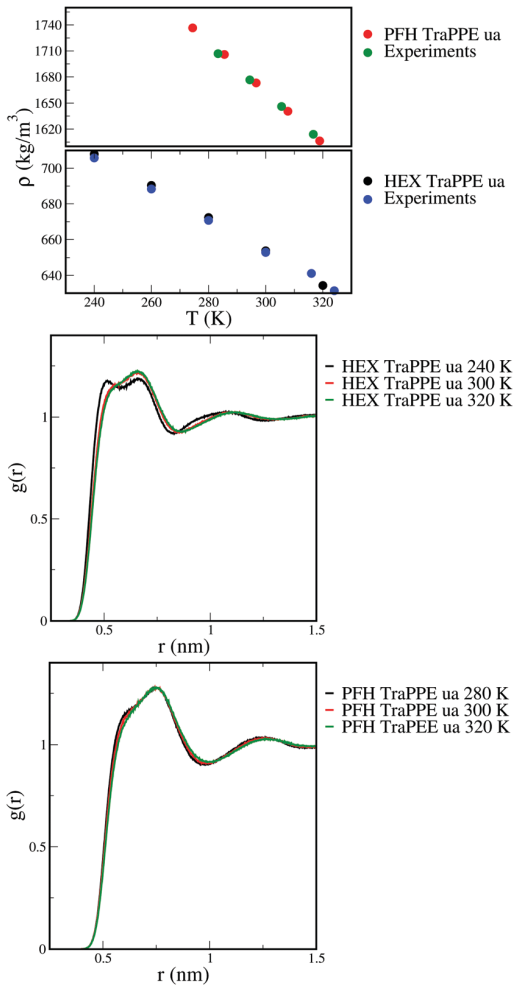


Fig. 2 Top panel: Comparison of bulk densities of PFH (red) with experimental data (green)[56] and for HEX (black) with experimental data (blue).[56] Central panel: RDF between the com of HEX molecules at different temperatures (black: 240 K, red 300 K and green: 320 K). Bottom panel: RDF between the com of PFH molecules at different temperatures (black: 280 K, red 300 K and green: 320 K).

lowest temperature of 280 K (black line). This shoulder flattens out with increasing temperature (300 K red line, 320 K green line).

4.2 IMC model for hexane

4.2.1 Coarse grained potentials. We derived a 3-bead CG model (see left hand side in Fig. 1) as described under Section 3.2.1. The bonded interactions were derived with BI (see eqn (2)) based on the probability distributions of the bond length between two adjacent A–B beads and on the distribution of the A–B–A angle. The corresponding potentials are shown in Fig. 3. As one sees in the top panel of Fig. 3, the bond potential exhibits two distinct minimum values. These correspond to the gauché (0.23 nm) and trans state (0.26 nm) of the HEX chain. Concisely the angular potential, which is presented in the bottom panel of Fig. 3, shows a minimum for the trans state (180°). The *cis*-state is not sampled by our CG model.

The non-bonded interactions were generated by application of the IMC method according to the procedure presented in Section 3.2.1. As shown in Fig. 4, the three different bead-bead target RDFs, A–A (FG: dashed red line; CG: solid black line), A–B (FG: dashed blue line; CG solid green line) and B–B (FG: dashed cyan line, CG: solid violet line), are well reproduced by our derived model. The complementary potentials show the strongest attraction between the A–A beads (dashed black line) and less attraction for the A–B (dashed green line) and B–B (dashed violet line) interaction. Further the interactions which involve the central B bead show a small potential barrier around 0.7 nm. This barrier is not present for the terminal A–A bead interaction. The small barrier and the lower attraction might be related to the central position of the B bead, which is surrounded by the slightly larger A beads. This sterically constrains a stronger and more direct interaction of the B beads with the surrounding non-bonded beads.

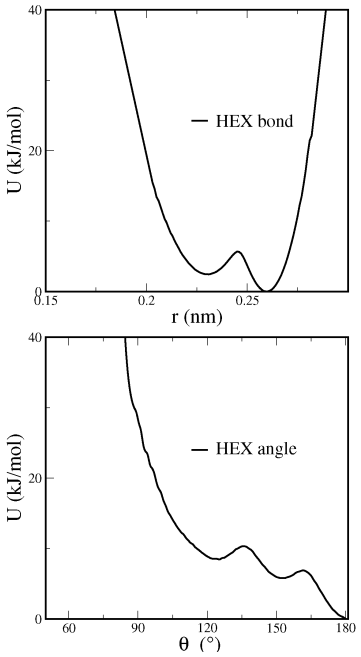


Fig. 3 Top panel: Bond potential to model the A–B bond obtained by BI. Bottom panel: Angular potential to model the A–B–A angle obtained by BI.

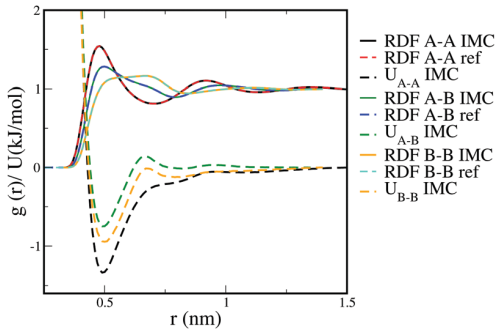


Fig. 4 Results of the IMC approach for HEX: final potentials for the A–A (dashed black line), A–B (dashed green line) and B–B (dashed violet line) interaction and their corresponding RDFs (solid lines with the same coloring as for the potentials) in comparison with the FG target RDFs: A–A: red dashed line, A–B: blue dashed line and B–B: cyan dashed line.

4.2.2 Evaluation of the pressure matching approach and assessment of the temperature transferability. We performed the DA-PM method in combination with our IMC model as described under Section 3.3. The coefficients, ψ_1 and ψ_2, were determined for each temperature separately by performing a rerun of the mapped FG ensemble with the CG FF derived at 300 K. We further applied the DN-PM approach for all temperatures individually. The latter was necessary since our CG model did reproduce neither the FG virial nor the volume fluctuations in combination with the DA-PM method as we will discuss below. In Fig. 5, we show that both coefficients appear to have an almost exact linear dependency on the temperature as indicated by the dashed lines, no matter if the DA-PM (top panel in Fig. 5) or the DN-PM (bottom panel in Fig. 5) is applied. Besides the linear trend, Fig. 5 reveals that the DN-PM approach leads to an increase in the absolute values for ψ_1 and to a decrease for ψ_2 at each temperature compared with the values obtained by the DA-PM method.

We tried to utilize the linear dependency shown in Fig. 5 to determine ψ_1 and ψ_2 at a state point of interest, in order to avoid the individual application of the DA-PM or DN-PM method at each state point. That is why we applied a LR between two chosen state points along the temperature range. The reference points for the different LR regimes tested are presented in Table 1. We used the equations of the regression lines to determine ψ_1 and ψ_2 at the remaining state points.

In Table 2 the values for ψ_1 and ψ_2 obtained by the different methods are presented and compared. As one sees, all DN-LR values for ψ_1 and ψ_2 which are comparable with the DN-PM approach. The coefficients obtained by the DA-PM approach lead to an underestimation of the average value of FG virial pressure by orders of magnitude as shown by the red circles in the top panel of Fig. 6. This further confirms the statement of DN that the coefficients obtained by the DA-PM need a self-consistent correction when coarse graining molecular liquids.[38] As shown in the top panel of Fig. 6 the average values of the virial pressure obtained by the DN-PM and DN-LR schemes are all closer to the ones of the FG model than the ones produced

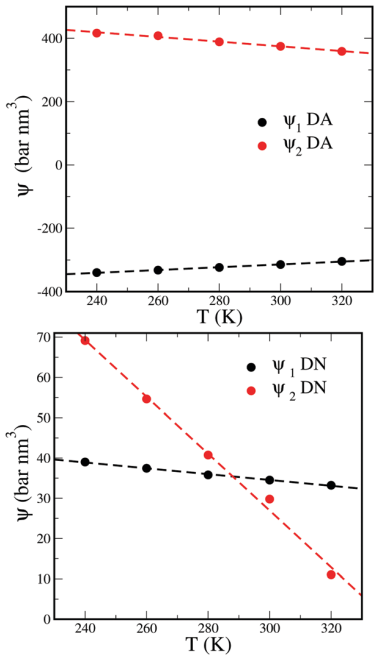


Fig. 5 Top panel: ψ_1 and ψ_2 obtained by DA-PM approach (ψ_1: black circles; ψ_2: red circles; dashed lines: linear fit). Bottom panel: ψ_1 and ψ_2 obtained by DN-PM approach (ψ_1: black circles; ψ_2: red circles; dashed lines: linear fit).

Table 1 Reference points for linear regression

Molecule	Label	T_1 (K)	T_2 (K)
HEX	DN-LR$_1$	240	300
	DN-LR$_2$	240	320
	DN-LR$_3$	260	300

by the DA model. To have a closer look on that, we present the average values of the virial pressure only for the DN-PM approach and the different LR schemes in comparison with the FG model in the bottom panel of Fig. 6. Despite the fact that all CG models reproduce the average value of the FG virial pressure with quantitative accuracy, they produce larger fluctuations. This is related to sampling, since our CG simulations were performed only for 10 ns. A 100 ns run showed that the average pressure at 300 K is about 0.8 bar and fluctuates for about 0.3 bar which is comparable to the FG simulations as shown on the first line in Table 3.

To analyze the thermal transferability of our CG models we computed the densities and the thermal expansion for all of the models and compared them with the FG results. In the top panel of Fig. 7, we show that the densities produced by the DA-PM approach (red circles) underestimate the bulk densities of the FG system (black circles). The DN-PM approach leads to an accurate reproduction of the FG densities at all DN-LR schemes tested. Likewise there is almost no difference between the different LR approaches. This means that we can use the LR scheme not only to interpolate between two states, but also to extrapolate. As a consequence, we basically can choose any two

Table 2 PM coefficients for HEX for the different methods tested and the number of iterations needed for the DN-PM approach to converge

T (K)	Method	ψ_1 (bar nm^3)	ψ_2 (bar nm^3)	Iterations
240	DA	−339.641	416.324	—
	DN	39.007	69.138	3
	DN-LR$_1$	—	—	—
	DN-LR$_2$	—	—	—
	DN-LR$_3$	38.903	67.027	
260	DA	−331.929	408.081	—
	DN	37.605	54.614	3
	DN-LR$_1$	37.509	56.028	—
	DN-LR$_2$	37.558	54.609	—
	DN-LR$_3$	—	—	—
280	DA	−323.544	388.691	—
	DN	35.806	40.748	3
	DN-LR$_1$	36.011	42.913	—
	DN-LR$_2$	36.109	40.081	—
	DN-LR$_3$	35.979	42.207	—
300	DA	−314.187	374.519	—
	DN	34.514	29.793	3
	DN-LR$_1$	—	—	—
	DN-LR$_2$	34.651	25.553	—
	DN-LR$_3$	—	—	—
320	DA	−304.237	358.606	—
	DN	33.213	11.027	3
	DN-LR$_1$	33.016	16.683	—
	DN-LR$_2$	—	—	—
	DN-LR$_3$	33.049	17.387	—

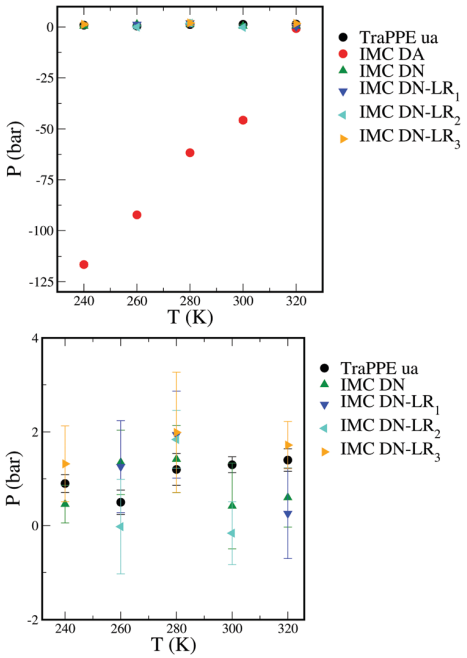


Fig. 6 Average values of the virial pressure for HEX at different temperatures: Top panel: Black circles TraPPE-ua, red circles: DA-PM, green upper triangles: DN-PM, blue lower triangles: DN-LR$_1$, cyan left triangles: DN-LR$_2$, orange right triangles: DN-LR$_3$. Bottom panel: Black circles TraPPE-ua green upper triangles: DN-PM, blue lower triangles: DN-LR$_1$, cyan left triangles: DN-LR$_2$, orange right triangles: DN-LR$_3$.

Table 3 Thermodynamic properties of HEX at 300 K for the FG system and all CG models: thermal expansion coefficient α, isothermal compressibility κ_T, bulk density ρ, and average virial pressure p

Method	α (K^{-1})	κ_T ($m^2\ N^{-1}$)	ρ (kg m^{-3})	p (bar)
TraPPE-ua	1.30×10^{-3}	2.25×10^{-9}	653.6 ± 0.06	1.3 ± 0.17
DN	1.33×10^{-3}	1.95×10^{-9}	653.6 ± 0.10	0.4 ± 0.92
DN-LR_2	1.33×10^{-3}	1.93×10^{-9}	653.7 ± 0.08	-0.2 ± 0.67

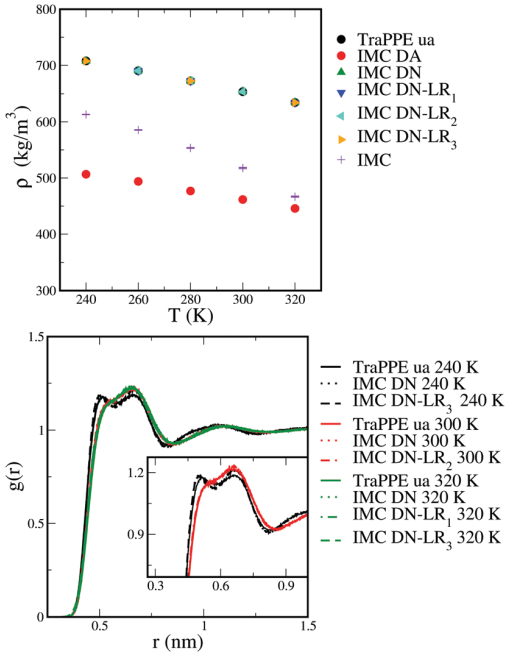


Fig. 7 Top panel: Average bulk densities of HEX at different temperatures: black circles TraPPE-ua, red circles: DA-PM, green upper triangles: DN-PM, blue lower triangles: DN-LR_1, cyan left triangles: DN-LR_2, orange right triangles: DN-LR_3, violet crosses: no correction. Bottom panel: Comparison of the com RDFs of HEX between the FG model (solid lines) the DN-PM model (dotted lines) and the DN-LR models (DN-LR_1: dashed, DN-LR_2: dashed-dotted and DN-LR_3: dotted-dashed) at 240 K (black), 300 K (red) and 320 K (green). The inset compares the HEX com RDF at 240 K and 300 K in a higher resolution.

temperatures and we do not even have to include the sate point at which the configurational part of the FF has been parametrized in order to get the correct thermal expansion behaviour. To further prove that the volume dependent potential is needed to get a quantitative reproduction of the FG system, we computed the densities produced by the CG model, derived at 300 K, at different temperatures without any correction. As the violet crosses in the top panel of Fig. 7 show, the densities are underestimated without the additional volume dependent potential while the thermal expansion (slope) is over-predicted.

To assess the impact of the volume dependent potential on the structure of the system, we also examined the RDF between the com of the HEX molecules. As shown in the bottom panel of Fig. 7 at 300 K (red curves) and at 320 K (green curves) our CG models, with bonded and non-bonded interactions derived at 300 K, reproduce the structure of the FG system with high accuracy (300 K: red solid line, 320 K: green solid line). At 240 K (black curves) the structural match is less accurate, but still acceptable, since the qualitative shape is still reproduced. The structural mismatch can be caused by the fact that the interactions derived at 300 K might not be good enough to accurately describe the structure at 240 K, since at this low temperature many-body effects could play a larger role than at 300 K. Nevertheless the density and the virial pressure are quantitatively reproduced, that is why the slight mismatch in the structure might not be too critical in order to describe the system at 240 K physically correct with our CG model. The good agreement for the higher temperatures might be related to the fact the systems are either exactly at or closer to the reference state point of 300 K. Though the structural mismatch decreases with increasing temperature, all CG models show a lower compressibility when compared to the FG system (see Table 3). This is highlighted by the inset in the bottom panel of Fig. 7. Here the com RDFs at 240 K (black curves) show a slight over-structuring of the system which result in lower compressibility. At 300 K (red curves in the inset) the over-structuring is less, but still occurs.

Despite this small loss in the isothermal compressibility and the accuracy in the com RDF, the DN-PM and DN-LR CG models provide an accurate reproduction of the thermal expansion behaviour of the FG system.

In Table 3 we present the results obtained from the 10 ns CG simulations for all models at 300 K in comparison with the FG reference.

Table 3 summarizes our main conclusions. It is possible to achieve temperature transferability with IMC based CG models as shown by the thermal expansion coefficient, α, obtained from the DN-PM and DN-LR approach. The approach presented here does not require a complete generation, *i.e.* bonded and non-bonded interactions, of a structure based CG model at two different state points as in the approach of Farah *et al.*[30] This is an improvement, since the computational cost is reduced and moreover we do not have to combine two approximate solutions of the inverse problem discussed in Section 2.1. Further we do not need a thermodynamic constraint, which adds a linear ramp to the non-bonded interactions of the system, in order to reproduce the pressure of the system and to enable simulations in the *NPT* ensemble.[5,22] This is proven by the average values of the virial pressure P and the bulk density ρ. It is also not necessary to compute averages over multiple states as performed by Moore *et al.* for systems of different densities.[29] Here we achieve the temperature transferability and *NPT* stability by the application of the DN-PM approach. This method enables an accurate reproduction of the FG virial pressure and the volume fluctuations of the FG system in the CG ensemble. We show that the set of coefficients, ψ_1 and ψ_2, determined by the DN-PM approach demonstrate a linear temperature dependency, which allows us to predict the coefficients *via* the DN-LR approach. This approach is a simple two state parametrization, which does not require any further information. Of course the improved temperature transferability and *NPT* stability comes with a cost. The cost of reduced accuracy in representing structural correlations and corresponding reduced accuracy in the isothermal compressibility, κ_T. We also note that the success of the DN-LR scheme does not

depend on the choice of the reference points, as shown in Fig. 6 and 7. Further, the state point, which was chosen during the parametrization of the configurational part of the CG FF, does not necessarily has to be taken as one of the reference points.

So far we have limited the temperature range to the stable liquid phase of hexane. To further test the predictive quality of our proposed linear regression approach and to explore the limitations of the model, we extended the temperature range from below the experimental freezing point of liquid HEX at 178 K to the experimental boiling point of 341 K. In the top panel of Fig. 8 the bulk densities for the new set of temperatures are presented. As one sees the global densities of the FG model (black circles) are still well reproduced, as exemplified by the DN-LR$_1$ CG model (red circles), even if we approach the freezing or the boiling point of hexane. Only at 100 K, which is far below the experimental freezing point, the bulk density of our CG model starts to diverge significantly from its reference. Despite the good representability of the bulk density, the local structure is not as accurately reproduced, as demonstrated in the bottom panel of Fig. 8. If the freezing point is approached (200 K, red curves) or even passed (100 K, black curves), the structural agreement between the model (dashed lines) and its reference (solid lines) gets less and less accurate. This might indicate that a simple pair potential, as the one obtained from IMC, might not be sufficient to accurately describe the local structure of a more ordered condensed phase. Oppositely if we approach the boiling point (340 K, green curves), the model (dashed line) accurately reproduces the structure of the reference system (solid line). Although our model lacks the ability to describe the structure of an ordered condensed phase, still it is possible to describe bulk properties of liquid hexane quite accurately over a broad range of temperatures without accounting for complex many-body contributions.

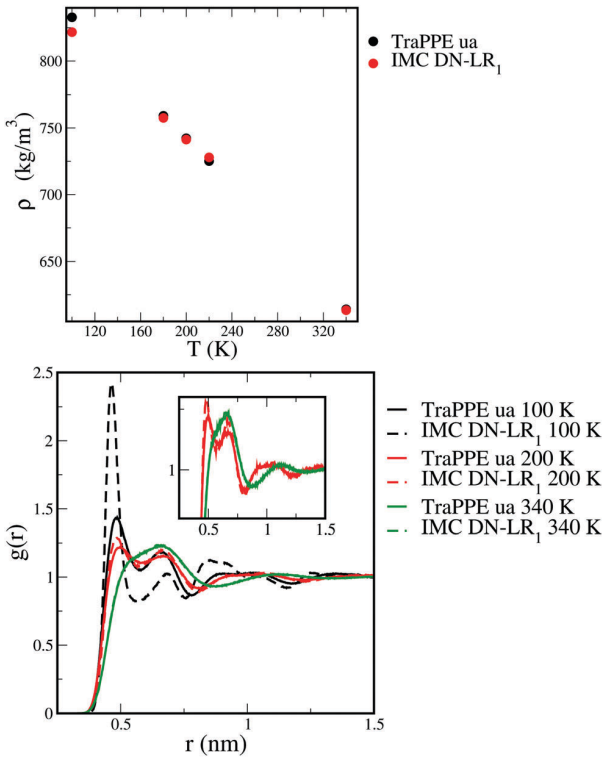


Fig. 8 Top panel: Average bulk densities of HEX at different temperatures: black circles TraPPE-ua, red circles: DN-LR$_1$. Bottom panel: Comparison of the com RDFs of HEX between the FG model (solid lines) the DN-LR$_1$ model (dashed lines) at 100 K (black), 200 K (red) and 340 K (green). The inset compares the HEX com RDF at 200 K and 340 K in a higher resolution.

4.3 Perfluorohexane

The previously discussed CG model for HEX has been derived on the basis of a relatively small degree of coarse graining (2 to 1 mapping scheme). To asses the impact of the mapping scheme and to further explore the DN-LR approach we derived an IMC model for PFH based on the 6 to 1 mapping scheme depicted on the right hand side in Fig. 1. Following the procedure reported under Section 3.2.1, we obtain the effective pair potential for the C–C interaction shown by the solid green line in Fig. 9. The corresponding RDF (solid black line in Fig. 9) matches the FG one (dashed red line). The final potential is weakly attractive and shows a wide potential well depth.

We again applied the DA-PM and DN-PM approach as described in Section 3.3, but this time only at 280 K and 300 K. We applied the DN-LR approach to obtain the coefficients at the remaining temperatures. The final values are presented in Table 4. As one sees, the application of the DN-PM method leads to a decrease in the values for both coefficients ψ_1 and ψ_2 compared to the DA-PM approach at 280 K and 300 K.

In the top panel of Fig. 10 we show that our DN-LR approach quite accurately reproduces the bulk densities of the underlying FG systems. Although a slight increase in the bulk density of the CG system is observed at 200 K and 340 K compared to the FG reference the quantitative agreement is still acceptable at both, higher and lower, temperatures. One should further note, if simulations are performed at 1 bar without the additional volume dependent potential term, the bulk density at 300 K would drop from 1673.9 kg m^{-3} to 13.5 kg m^{-3}. This may be avoided by setting the barostat pressure equal to the pressure which the CG model produces under *NVT* conditions at the

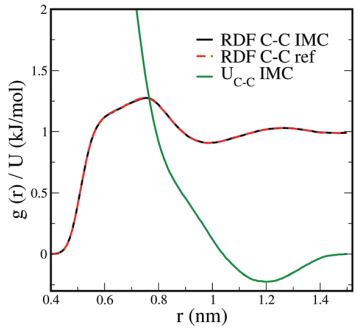


Fig. 9 Results of the IMC parametrization for PFH: IMC potential (solid green line) the target RDF (dashed red line) and the CG RDF (solid black line).

Table 4 PM coefficients for PFH for the different methods tested and the number of iterations needed for the DN-PM approach to converge

T (K)	Method	ψ_1 (bar nm^3)	ψ_2 (bar nm^3)	Iterations
200	DN-LR	343.470	243.080	—
220	DN-LR	335.020	183.786	—
240	DN-LR	326.570	124.492	—
260	DN-LR	318.120	65.198	—
280	DA	326.490	24.080	—
	DN	309.673	5.902	1
290	DN-LR	305.455	−23.743	—
300	DA	315.573	−26.827	—
	DN	301.223	−53.392	1
310	DN-LR	296.995	−83.037	—
320	DN-LR	292.770	−112.684	—
340	DN-LR	284.320	−171.978	—

density of the united atom model at 1 bar and 300 K (884.7 bar). However, then, no temperature transferability is achieved as indicated by the blue circles in the top panel of Fig. 10. In the bottom panel of Fig. 10 we show the comparison of the com RDF of the FG model (solid lines) with the CG model (dotted lines) at 200 K (black), 240 K (red), 300 K (green) and 340 K (blue). Comparable to HEX, the CG model cannot quantitatively

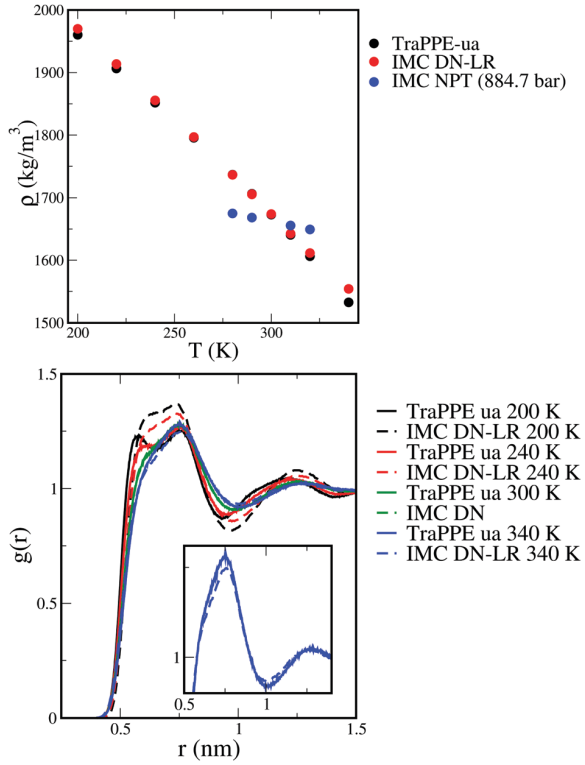


Fig. 10 Top panel: Average bulk densities of PFH at different temperatures: obtained from the TraPPE-ua FF (black circles), the DN-LR model (red circles), and from *NPT* simulations of the IMC model without the volume dependent potential at 884.7 bar. Bottom panel: Comparison of the com RDFs for PFH between the TraPEE-ua model (solid lines) and the DN-LR model (dotted lines) at 200 K (black), 240 K (red), 300 K (green) and 340 K. The inset compares the PFH com RDF at 340 K in a higher resolution.

Table 5 Thermodynamic properties of PFH at 300 K: thermal expansion coefficient α, isothermal compressibility κ_T, bulk density ρ, and average virial pressure p

Method	α (K^{-1})	κ_T (m^2 N^{-1})	ρ (kg m^{-3})	p (bar)
TraPPE-ua	1.52×10^{-3}	5.81×10^{-9}	1673.0 ± 0.17	1.1 ± 0.03
DN-LR	1.51×10^{-3}	5.76×10^{-9}	1673.9 ± 0.06	1.0 ± 0.01

reproduce the FG structure at lower temperatures. At 200 K and at 240 K the reference model shows two distinct peaks, which are not captured in the CG resolution. Instead, we observe a broad first peak with larger RDF values. The RDFs at lower temperatures resemble the shape of the RDF at 300 K (green curves), which was the reference temperature in the IMC process. This indicates that our effective pair potential cannot capture the structural features of a more ordered condensed phase, what leads to lower compressibility and a slight over-structuring of the system compared to the FG reference. The inset in the bottom panel of Fig. 10 reveals that close to the experimental boiling point at 340 K (blue curves), the RDF of the CG model (dotted blue line) tends to be a bit lower around the maximum value at around 0.75 nm compared to the FG model (solid blue line). This corresponds to the slightly larger bulk density the CG model produces at this temperature.

Finally in Table 5 we compare the thermodynamic properties of PFH at 300 K obtained by the FG model and our derived CG model. The generated model reproduces all thermodynamic bulk properties of the FG model quite accurately. Although a mismatch in the local structure is observed, the good representability and transferability of bulk properties emphasizes that our proposed DN-LR approach allows us to derive temperature transferable CG models, which do not require more than a configurational FF derived at one state point and the application of the DN-PM at two state points.

5 Conclusion and outlook

In this work we proposed a rather simple method to achieve temperature transferability of bulk properties for CG models of HEX and PFH derived with IMC. The proposed approach leaves the EPP untouched after they have been successfully derived, but requires further the application of the DN-PM approach at two state points. As we have demonstrated, these two state points can be chosen quite generally. The PM coefficients, ψ_1 and ψ_2, at the remaining state points of interest can be obtained by a simple LR approach. The generated models reproduce the average virial pressure and the thermal expansion coefficients of their underlying systems quite accurately, although only two state points are included in the parametrization of the model. This emphasizes the predictive quality of our proposed DN-LR approach. Nevertheless, the improvement of the thermal bulk behaviour comes with the cost of a loss of accuracy in the isothermal compressibility and a loss of accuracy in the representability of the local structure of the system. Especially if the system gets cooled down and the structure is determined more by long range interactions.

Despite the differences in representability and transferability between bulk and local properties, the method proposed can be seen as new starting point to develop good representable and transferable structure based CG models. We showed that it is possible to combine an iterative inverse structure based coarse graining method with a variational principle to improve the representability as well as the temperature transferability.

Nevertheless, in this study we were only focused on two liquid alkanes in the bulk phase. Future work might address the question if it is possible to achieve temperature transferability for more complex systems like mixtures, polymers in solutions or ionic liquids. Further the LR approach can also be tested in combination with other bottom coarse-graining methods like the MS-CG method,[10] the relative entropy method[57] or the conditional reversible work method.[58] It might also not be enough to assume simple pair wise additive non-bonded interactions for the CG system. As work by Villa *et al.*[59] or more recent by Sanyal and Shell[41] and DeLyser and Noid[40] suggests. Systems which undergo a phase transition or which are inhomogeneous, cannot be accurately described with EPP. There an extension to local density dependent potentials is necessary. Another open question is, if the additional volume dependent potential improves the representability of concentration dependent properties as well. KBIs, for example, have been accurately reproduced by the Kirkwood–Buff version of the IBI approach.[28] This method further enable transferability over a certain range of concentrations. In the context of PM, this has been addressed with an extended ensemble approach by DN so far.[39]

Conflicts of interest

There are no conflicts of interest.

Acknowledgements

The authors thank William Noid, Nicholas Dunn and Joseph Rudzinski for helpful comments regarding the application and implementation of the pressure matching approach. D. R. is thankful to M. Scott Shell for useful discussions. Funding has been granted by the Deutsche Forschungsgemeinschaft through the Collaborative Research Center SFB TRR146 Multiscale Simulation Methods for Soft Matter Systems.

References

1 C. Peter and K. Kremer, *Faraday Discuss.*, 2010, **144**, 9–24.
2 E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero and N. F. A. van der Vegt, *Soft Matter*, 2013, **9**, 2108–2119.
3 W. G. Noid, *J. Chem. Phys.*, 2013, **139**, 090901.
4 W. Schommers, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1983, **28**, 3599–3605.
5 D. Reith, M. Pütz and F. Müller-Plathe, *J. Comput. Chem.*, 2003, **24**, 1624–1636.
6 A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1995, **52**, 3730–3737.
7 J. W. Mullinax and W. G. Noid, *Phys. Rev. Lett.*, 2009, **103**, 198104.
8 J. W. Mullinax and W. G. Noid, *J. Phys. Chem. C*, 2010, **114**, 5661–5674.
9 F. Ercolessi and J. B. Adams, *Europhys. Lett.*, 1994, **26**, 583–588.
10 S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469–2473.
11 W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das and H. C. Andersen, *J. Chem. Phys.*, 2008, **128**, 244114.
12 W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen and G. A. Voth, *J. Chem. Phys.*, 2008, **128**, 244115.
13 D. Ivanizki, PhD thesis, Johannes Gutenberg Universität Mainz, 2015.
14 A. Lyubartsev, A. Mirzoev, L. Chen and A. Laaksonen, *Faraday Discuss.*, 2010, **144**, 43–56.
15 D. Rosenberger, M. Hanke and N. F. A. van der Vegt, *Eur. Phys. J.: Spec. Top.*, 2016, **225**, 1323–1345.
16 J. F. Rudzinski and W. G. Noid, *Eur. Phys. J.: Spec. Top.*, 2015, **224**, 2193–2216.
17 J.-P. Hansen and I. R. McDonald, *Theory of simple liquids*, Elsevier/Academic Press, Amsterdam, Boston, 3rd edn, 2007.
18 C. Russ, M. Brunner, C. Bechinger and H. H. von Grünberg, *EPL*, 2005, **69**, 468.
19 J. F. Rudzinski and W. G. Noid, *J. Phys. Chem. B*, 2014, **118**, 8295–8312.
20 S. Riniker, J. R. Allison and W. F. van Gunsteren, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12423.
21 M. Guenza, *Eur. Phys. J.: Spec. Top.*, 2015, **224**, 2177–2191.
22 H. Wang, C. Junghans and K. Kremer, *Eur. Phys. J. E: Soft Matter Biol. Phys.*, 2009, **28**, 221–229.
23 J. Ghosh and R. Faller, *Mol. Simul.*, 2007, **33**, 759–767.
24 E. C. Allen and G. C. Rutledge, *J. Chem. Phys.*, 2008, **128**, 154115.
25 J. W. Mullinax and W. G. Noid, *J. Chem. Phys.*, 2009, **131**, 104110.
26 E. Brini and N. F. A. van der Vegt, *J. Chem. Phys.*, 2012, **137**, 154113.
27 P. Ganguly and N. F. A. van der Vegt, *J. Chem. Theory Comput.*, 2013, **9**, 5247–5256.
28 P. Ganguly, D. Mukherji, C. Junghans and N. F. A. van der Vegt, *J. Chem. Theory Comput.*, 2012, **8**, 1802–1807.
29 T. C. Moore, C. R. Iacovella and C. McCabe, *J. Chem. Phys.*, 2014, **140**, 224104.
30 K. Farah, A. C. Fogarty, M. C. Böhm and F. Müller-Plathe, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2894–2902.
31 T. Murtola, E. Falck, M. Karttunen and I. Vattulainen, *J. Chem. Phys.*, 2007, **126**, 075101.
32 N. G. Almarza and E. Lomba, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2003, **68**, 011202.
33 A. Das and H. C. Andersen, *J. Chem. Phys.*, 2010, **132**, 164106.
34 N. Ashcroft and D. Stroud, in *N. Ashcroft and D. Stroud*, ed. H. Ehrenreich, F. Seitz and D. Turnbull, Academic Press, 1978, vol. 33, pp. 1–81.
35 E. Canessa, M. Silbert and M. Grimson, *Mol. Phys.*, 1988, **64**, 1195–1201.
36 G. C. Barker, M. J. Grimson and M. Silbert, *Mol. Phys.*, 1995, **84**, 211–215.

37 C. N. Likos, *Phys. Rep.*, 2001, **348**, 267–439.
38 N. J. H. Dunn and W. G. Noid, *J. Chem. Phys.*, 2015, **143**, 243148.
39 N. J. H. Dunn and W. G. Noid, *J. Chem. Phys.*, 2016, **144**, 204124.
40 M. R. DeLyser and W. G. Noid, *J. Chem. Phys.*, 2017, **147**, 134111.
41 T. Sanyal and M. S. Shell, *J. Chem. Phys.*, 2016, **145**, 034109.
42 N. F. A. van der Vegt, C. Peter and K. Kremer, in *Coarse-Graining of Condensed Phase and Biomolecular Systems*, ed. G. A. Voth, Chapman and Hall/CRC Press, Taylor Francis Group, 2008, ch. 25, pp. 379–397.
43 V. Rühle, C. Junghans, A. Lukyanov, K. Kremer and D. Andrienko, *J. Chem. Theory Comput.*, 2009, **5**, 3211–3223.
44 N. J. H. Dunn, T. T. Foley and W. G. Noid, *Acc. Chem. Res.*, 2016, **49**, 2832–2840.
45 D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
46 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.
47 L. Zhang and J. I. Siepmann, *J. Phys. Chem. B*, 2005, **109**, 2911–2919.
48 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
49 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
50 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
51 S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle and C. Junghans, *PLoS One*, 2015, **10**, e0131754.
52 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.
53 N. J. H. Dunn, K. M. Lebold, M. R. DeLyser, J. F. Rudzinski and W. Noid, *J. Phys. Chem. B*, 2017, DOI: 10.1021/acs.jpcb.7b09993.
54 G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.
55 G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, *Mol. Phys.*, 1996, **87**, 1117–1157.
56 B. D. Smith and R. Srivastava, *Thermodynamic data for pure compounds*, Elsevier; Distributors for the US and Canada, Elsevier Science Pub. Co, Amsterdam, New York, New York, NY, USA, 1986.
57 M. S. Shell, *J. Chem. Phys.*, 2008, **129**, 144108.
58 E. Brini, V. Marcon and N. F. A. van der Vegt, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10468.
59 A. Villa, C. Peter and N. F. A. van der Vegt, *J. Chem. Theory Comput.*, 2010, **6**, 2434–2444.

Conclusion and Outlook

In this thesis new insights into the field of structure-based bottom-up coarse graining are provided. By comparing the IMC approach with different IBI-type methods, we show that IMC converges much faster to a solution. The reason for this is that IMC is an exact Newton method to solve inverse problems. But, IMC suffers from numerical instabilities, which require additional regularization. Further, we demonstrate that faster convergence and the more profound theoretical background does not suffice to overcome the main representability problem in structure based coarse graining. This means that IMC models cannot quantitatively reproduce any other quantity than the pair structure. This indicates that reproducing structure might be a crude target to estimate the total probability distribution of a configuration space.

Nevertheless, we showed it is possible to overcome these shortcomings by adding additional energetic contributions to the system's Hamiltonian. In particular we showed that additional volume dependent potentials, for which we derived a novel and computationally cheap parametrization, improve the representability of thermodynmaic properties of IMC models for liquid alkanes for different mapping schemes. What is more, the additional energetic contribution improved the temperature transferability in the liquid state.

Moreover we investigated the effect of local density dependent potentials on the concentration transferability of implicit solvent models for homogeneous mixtures of methanol and water. Here, we showed that local density potentials improved the transferability and representability of CG models only in the direction of decreasing methanol concentration and only for water in implicit methanol. Thus, to capture structural changes, which come along with increasing concentration of methanol, additional degrees of freedom have to be included in the model, which allow for example for the formation of hydrogen bonds. This illustrates the power of bottom-up coarse graining to identify, which degrees of freedom are necessary for an accurate description for a model at a CG level. Additionally, we also prove numerically and analytically that IMC and relative entropy optimization can yield the same set of effective interaction potentials. This shows that it is possible to derive structure-based CG models without explicitly accounting for the structure in the parametrization procedure.

Further, we showed that the concentration transferability of IMC models with respect to the structure can be improved by systematically selecting a state point for the model parametrization. This ideal state point could be identified by computing the relative entropies for different IMC models at different concentrations. This approach is the first of its kind and requires more work in the future as many issues have to be solved. It has to be tested if the findings also hold for other bottom-up coarse graining methods than IMC. Further, we could only identify an ideal state point for a simple model system. So future work should target to compute relative entropies for more applied systems than mixtures of particles, which interact with 6-12 Lennard-Jones potentials. The reason we think it is worth to pursue that direction is that we observed trends in concentration transferability for the model system as well as for binary mixtures of liquid alkanes. The long term goal is to combine the extended Hamiltonian methods, which are necessary to overcome the poor sampling of the full configuration space with IMC models, with the parametrization at an ideal state point. This should lead to representable and transferable structure based CG models.

Another aspect which carefully has to be analyzed in the future is the role of the mapping scheme. There is still no systematic way how to select it. But, as work by Foley et al. suggests there seems to be an optimal mapping scheme.[41] The problem is that in order to identify this mapping scheme one has to be able to exactly compute the m-PMF. Yet, this is not possible for most systems. Nevertheless, the

effect of different mapping schemes on the performance of CG models and the interplay with a chosen coarse graining method can give meaningful insights into the performance of a CG model.[42,94] Instead of focusing only on the optimal resolution, one could also choose different reference points to determine the position of a CG particle. In recent work the optimal position of a CG particle is determined based on graph-theoretic principles, and not solely on the basis of the center of mass coordinates.[155] Another possibility to improve the choice of a mapping is to select it on the fly during the process of coarse graining as recently proposed.[43,156] All these novel mapping approaches might significantly improve the quality of a CG model.

But, an ideal mapping cannot compensate for all methodological shortcomings. Thus, novel approaches are required to generate the interaction potentials for CG models. A tempting approach is to combine bottom-up methods with top-down methods. Here, an experimental bias is introduced on the reference model for the bottom-up process, which as an example leads to more accurate CG models for carbohydrates.[157] Another idea is to directly constrain the algorithm to simultaneously reproduce multiple target properties as structure and surface tension[136], structure and pressure[137] or structure and dipole moments.[158] These constrained algorithm can further be combined with the extended Hamiltonian approaches discussed in chapter 3.2, 3.3 and 3.4.[147]

Additionally, current approaches also couple multiple CG force fields together. Based on a local order parameter stochastic transitions between these different force fields occur.[149,159] As a consequence a CG model is described as a combination of multiple force fields. This has lead to CG model which can explicitly reproduce Hydrogen bonding in several non-polar and polar liquids[153] and more recently for a consistent description of the long-time dynamics of the tetraalanine peptide in water.[160]

Nevertheless, all these approaches still require individual parametrization. Thus, one can consider to combine the rather brute-force approaches with artificial neural networks or other machine learning approaches to speed up the parametrization process.[161,162] Machine learning approaches may not only speed-up the parametrization of a CG model, they also enable an extension of the CG model beyond pairwise additivity.[163] Hence, a more accurate approximation to the m-PMF is possible.

A highly accurate CG model should also quantitatively describe dynamic properties. Although not part of this thesis, we want to briefly describe how to build dynamically consistent CG models, as it completes the discussion on bottom-up coarse graining. By reducing the number of degrees of freedom and smoothening the potential energy surface CG models get accelerated.[5,164–166] Thus, their dynamics are wrong, which prohibits the study of dynamic or transport properties. In some cases simple rescaling algorithm can restore the correct long time dynamics, but these approaches are constraint especially to single component systems.[167,168] A more rigorous way to construct dynamically consistent CG models is to use the Mori-Zwanzig (MZ) projection[169,170] to derive the generalized Langevin equation (GLE), a special type of equation of motion.[166] Based on the assumption of complete time-scale separation, which means that fluctuating forces equilibrate faster than the center of motion of CG particles, one can derive the dissipative particle dynamics (MZ-DPD) equations of motion from the GLE.[14,15,171,172] MZ-DPD has been successfully applied to derive CG models for Lennard-Jones fluids,[173,174] repulsive star polymers[166,175–177] and liquids[178–181] at low densities. More recently, MZ-DPD has also be applied to describe systems at more realistic conditions.[182,183]

Another way to restore accurate dynamics on a CG level is to construct so called Markov state models (MSM).[184] MSM are CG kinetic models of FG trajectories. Here, a FG trajectory gets split into different microstates and based on the transition rate between these different microstates, information on the dynamics can be obtained. MSM have been used to restore the dynamic behavior for CG models of peptides.[185,186]

All these aspects, mapping, methods and dynamics, provide a large road map to build accurate CG mod-

els. It is up to the user which route to follow. On the other hand, one could also use all these tools to better understand what happens at a FG level. This thesis provides new pathways to follow both directions.

References

[1] C. N. Likos, "Effective interactions in soft condensed matter physics," *Physics Reports*, vol. 348, pp. 267–439, July 2001.

[2] C. Peter and K. Kremer, "Multiscale simulation of soft matter systems," *Faraday Discuss.*, vol. 144, pp. 9–24, 2010.

[3] E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodríguez-Ropero, and N. F. A. van der Vegt, "Systematic coarse-graining methods for soft matter simulations âĂŞ a review," *Soft Matter*, vol. 9, no. 7, pp. 2108–2119, 2013.

[4] A. Kovalenko and S. Gusarov, "Multiscale methods framework: self-consistent coupling of molecular theory of solvation with quantum chemistry, molecular simulations, and dissipative particle dynamics," *Physical Chemistry Chemical Physics*, vol. 20, no. 5, pp. 2947–2969, 2018.

[5] W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, "Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates," *Acta Polymerica*, vol. 49, pp. 61–74, Feb. 1998.

[6] M. Murat and K. Kremer, "From many monomers to many polymers: Soft ellipsoid model for polymer melts and mixtures," *The Journal of Chemical Physics*, vol. 108, pp. 4340–4348, Mar. 1998.

[7] F. Müller-Plathe, "Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back," *ChemPhysChem*, vol. 3, pp. 754–769, Sept. 2002.

[8] M. Praprotnik, L. D. Site, and K. Kremer, "Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution," *Annual Review of Physical Chemistry*, vol. 59, pp. 545–571, May 2008.

[9] C. Peter and K. Kremer, "Multiscale simulation of soft matter systems âĂŞ from the atomistic to the coarse-grained level and back," *Soft Matter*, vol. 5, no. 22, p. 4357, 2009.

[10] B. J. Alder and T. E. Wainwright, "Studies in Molecular Dynamics. I. General Method," *The Journal of Chemical Physics*, vol. 31, pp. 459–466, Aug. 1959.

[11] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *The Journal of Chemical Physics*, vol. 21, pp. 1087–1092, June 1953.

[12] A. Rahman, "Correlations in the Motion of Atoms in Liquid Argon," *Physical Review*, vol. 136, pp. A405–A411, Oct. 1964.

[13] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen, "Multiscale modeling of emergent materials: biological and soft matter," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, p. 1869, 2009.

[14] P. J. Hoogerbrugge and J. M. V. A. Koelman, "Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics," *Europhysics Letters (EPL)*, vol. 19, pp. 155–160, June 1992.

[15] P. Español and P. Warren, "Statistical Mechanics of Dissipative Particle Dynamics," *Europhysics Letters (EPL)*, vol. 30, pp. 191–196, May 1995.

[16] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," *Journal of Molecular Biology*, vol. 103, pp. 227–249, May 1976.

[17] J. Gao, "Methods and applications of combined quantum mechanical and molecular mechanical potentials," *Reviews in computational chemistry*, vol. 7, pp. 119–186, 1996.

[18] M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, "ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(*t* -Bu) $_3$) $_2$ + H $_2$ Oxidative Addition," *The Journal of Physical Chemistry*, vol. 100, pp. 19357–19363, Jan. 1996.

[19] P. Carloni, U. Rothlisberger, and M. Parrinello, "The Role and Perspective of Ab Initio Molecular Dynamics in the Study of Biological Systems," *Accounts of Chemical Research*, vol. 35, pp. 455–464, June 2002.

[20] R. E. Bulo, B. Ensing, J. Sikkema, and L. Visscher, "Toward a Practical Method for Adaptive QM/MM Simulations," *Journal of Chemical Theory and Computation*, vol. 5, pp. 2212–2221, Sept. 2009.

[21] K. Kreis, M. E. Tuckerman, D. Donadio, K. Kremer, and R. Potestio, "From Classical to Quantum and Back: A Hamiltonian Scheme for Adaptive Multiresolution Classical/Path-Integral Simulations," *Journal of Chemical Theory and Computation*, vol. 12, pp. 3030–3039, July 2016.

[22] M. Praprotnik, L. Delle Site, and K. Kremer, "Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly," *The Journal of Chemical Physics*, vol. 123, p. 224106, Dec. 2005.

[23] M. Praprotnik, L. Delle Site, and K. Kremer, "Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids," *Physical Review E*, vol. 73, June 2006.

[24] M. Praprotnik, L. Delle Site, and K. Kremer, "A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation," *The Journal of Chemical Physics*, vol. 126, p. 134902, Apr. 2007.

[25] R. Potestio, C. Peter, and K. Kremer, "Computer Simulations of Soft Matter: Linking the Scales," *Entropy*, vol. 16, pp. 4199–4245, July 2014.

[26] P. Español and M. Revenga, "Smoothed dissipative particle dynamics," *Physical Review E*, vol. 67, Feb. 2003.

[27] P. M. Kulkarni, C.-C. Fu, M. S. Shell, and L. Gary Leal, "Multiscale modeling with smoothed dissipative particle dynamics," *The Journal of Chemical Physics*, vol. 138, p. 234105, June 2013.

[28] N. D. Petsev, L. G. Leal, and M. S. Shell, "Hybrid molecular-continuum simulations using smoothed dissipative particle dynamics," *The Journal of Chemical Physics*, vol. 142, p. 044101, Jan. 2015.

[29] N. D. Petsev, L. G. Leal, and M. S. Shell, "Multiscale simulation of ideal mixtures using smoothed dissipative particle dynamics," *The Journal of Chemical Physics*, vol. 144, p. 084115, Feb. 2016.

[30] N. D. Petsev, L. G. Leal, and M. S. Shell, "Coupling discrete and continuum concentration particle models for multiscale and hybrid molecular-continuum simulations," *The Journal of Chemical Physics*, vol. 147, p. 234112, Dec. 2017.

[31] L. B. Lucy, "A numerical approach to the testing of the fission hypothesis," *The Astronomical Journal*, vol. 82, p. 1013, Dec. 1977.

[32] J. J. Monaghan, "Smoothed Particle Hydrodynamics," *Annual Review of Astronomy and Astrophysics*, vol. 30, pp. 543–574, Sept. 1992.

[33] M. Müller and G. D. Smith, "Phase separation in binary mixtures containing polymers: A quantitative comparison of single-chain-in-mean-field simulations and computer simulations of the corresponding multichain systems: Phase Separation in Binary Mixtures," *Journal of Polymer Science Part B: Polymer Physics*, vol. 43, pp. 934–958, Apr. 2005.

[34] K. C. Daoulas, M. Müller, J. J. de Pablo, P. F. Nealey, and G. D. Smith, "Morphology of multicomponent polymer systems: single chain in mean field simulation studies," *Soft Matter*, vol. 2, no. 7, pp. 573–583, 2006.

[35] S. L. Bore, G. Milano, and M. Cascella, "Hybrid Particle-Field Model for Conformational Dynamics of Peptide Chains," *Journal of Chemical Theory and Computation*, vol. 14, pp. 1120–1130, Feb. 2018.

[36] P. Ahlrichs and B. Dünweg, "Simulation of a single polymer chain in solution by combining lattice Boltzmann and molecular dynamics," *The Journal of Chemical Physics*, vol. 111, pp. 8225–8239, Nov. 1999.

[37] C. K. Aidun, Y. Lu, and E.-J. Ding, "Direct analysis of particulate suspensions with inertia using the discrete Boltzmann equation," *Journal of Fluid Mechanics*, vol. 373, pp. 287–311, Oct. 1998.

[38] O. Berk Usta, A. J. C. Ladd, and J. E. Butler, "Lattice-Boltzmann simulations of the dynamics of polymer solutions in periodic and confined geometries," *The Journal of Chemical Physics*, vol. 122, p. 094902, Mar. 2005.

[39] V. Boţan, V. D. Ustach, K. Leonhard, and R. Faller, "Development and Application of a Coarse-Grained Model for PNIPAM by Iterative Boltzmann Inversion and Its Combination with Lattice Boltzmann Hydrodynamics," *The Journal of Physical Chemistry B*, vol. 121, pp. 10394–10406, Nov. 2017.

[40] W. G. Noid, "Perspective: Coarse-grained models for biomolecular systems," *The Journal of Chemical Physics*, vol. 139, p. 090901, Sept. 2013.

[41] T. T. Foley, M. S. Shell, and W. G. Noid, "The impact of resolution upon entropy and information in coarse-grained models," *The Journal of Chemical Physics*, vol. 143, p. 243104, Dec. 2015.

[42] M. Dallavalle and N. F. A. van der Vegt, "Evaluation of mapping schemes for systematic coarse graining of higher alkanes," *Physical Chemistry Chemical Physics*, vol. 19, no. 34, pp. 23034–23042, 2017.

[43] Y. Han, J. F. Dama, and G. A. Voth, "Mesoscopic coarse-grained representations of fluids rigorously derived from atomistic models," *The Journal of Chemical Physics*, vol. 149, p. 044104, July 2018.

[44] G. A. Voth, ed., *Coarse-graining of condensed phase and biomolecular systems*. Boca Raton: CRC Press, 2009. OCLC: 232636723.

[45] M. Deserno, "Mesoscopic Membrane Physics: Concepts, Simulations, and Selected Applications," *Macromolecular Rapid Communications*, vol. 30, pp. 752–771, May 2009.

[46] F. Schmid, "Toy amphiphiles on the computer: What can we learn from generic models?," *Macromolecular Rapid Communications*, vol. 30, pp. 741–751, May 2009.

[47] F. A. Detcheverry, D. Q. Pike, U. Nagpal, P. F. Nealey, and J. J. de Pablo, "Theoretically informed coarse grain simulations of block copolymer melts: method and applications," *Soft Matter*, vol. 5, no. 24, p. 4858, 2009.

[48] D. Q. Pike, F. A. Detcheverry, M. Müller, and J. J. de Pablo, "Theoretically informed coarse grain simulations of polymeric systems," *The Journal of Chemical Physics*, vol. 131, p. 084903, Aug. 2009.

[49] M. Hömberg and M. Müller, "Main phase transition in lipid bilayers: Phase coexistence and line tension in a soft, solvent-free, coarse-grained model," *The Journal of Chemical Physics*, vol. 132, p. 155104, Apr. 2010.

[50] J. D. Honeycutt and D. Thirumalai, "Metastability of the folded states of globular proteins," *Proc Natl Acad Sci USA*, vol. 87, p. 3526, May 1990.

[51] I. R. Cooke, K. Kremer, and M. Deserno, "Tunable generic model for fluid bilayer membranes," *Physical Review E*, vol. 72, July 2005.

[52] G. Gompper, M. Schick, C. Domb, M. S. Green, J. L. Lebowitz, and G. Gompper, *Self-assembling amphiphilic systems*. No. ed. by C. Domb ... ; Vol. 16 in Phase transitions and critical phenomena, London: Acad. Press, 2. printing ed., 1995. OCLC: 258008528.

[53] G. Ciccotti, M. Ferrario, and K. Binder, "Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 2," 2006. OCLC: 758353671.

[54] S. Komura, "Mesoscale structures in microemulsions," *Journal of Physics: Condensed Matter*, vol. 19, p. 463101, Nov. 2007.

[55] M. Venturoli, M. Maddalenasperotto, M. Kranenburg, and B. Smit, "Mesoscopic models of biological membranes," *Physics Reports*, vol. 437, pp. 1–54, Dec. 2006.

[56] G. Brannigan, L. C.-L. Lin, and F. L. H. Brown, "Implicit solvent simulation models for biomembranes," *European Biophysics Journal*, vol. 35, pp. 104–124, Jan. 2006.

[57] J. C. Shillcock and R. Lipowsky, "The computational route from bilayer membranes to vesicle fusion," *Journal of Physics: Condensed Matter*, vol. 18, pp. S1191–S1219, July 2006.

[58] K. Kremer and G. S. Grest, "Dynamics of entangled linear polymer melts: A molecularâĂŘdynamics simulation," *The Journal of Chemical Physics*, vol. 92, pp. 5057–5086, Apr. 1990.

[59] R. Zangi, R. Zhou, and B. J. Berne, "Urea's Action on Hydrophobic Interactions," *Journal of the American Chemical Society*, vol. 131, pp. 1535–1541, Feb. 2009.

[60] D. Nayar and N. F. A. van der Vegt, "Cosolvent Effects on Polymer Hydration Drive Hydrophobic Collapse," *The Journal of Physical Chemistry B*, vol. 122, pp. 3587–3595, Apr. 2018.

[61] S. J. Marrink, A. H. de Vries, and A. E. Mark, "Coarse Grained Model for Semiquantitative Lipid Simulations," *The Journal of Physical Chemistry B*, vol. 108, pp. 750–760, Jan. 2004.

[62] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," *The Journal of Physical Chemistry B*, vol. 111, pp. 7812–7824, July 2007.

[63] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, "The MARTINI Coarse-Grained Force Field: Extension to Proteins," *Journal of Chemical Theory and Computation*, vol. 4, pp. 819–834, May 2008.

[64] S. J. Marrink and D. P. Tieleman, "Perspective on the Martini model," *Chemical Society Reviews*, vol. 42, no. 16, p. 6801, 2013.

[65] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, "A Coarse Grain Model for Phospholipid Simulations," *The Journal of Physical Chemistry B*, vol. 105, pp. 4464–4470, May 2001.

[66] W. Shinoda, R. DeVane, and M. L. Klein, "Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants," *Molecular Simulation*, vol. 33, pp. 27–36, Jan. 2007.

[67] C. Tanford, "Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions," *Journal of the American Chemical Society*, vol. 86, pp. 2050–2059, May 1964.

[68] B. Moeser and D. Horinek, "Unified Description of Urea Denaturation: Backbone and Side Chains Contribute Equally in the Transfer Model," *The Journal of Physical Chemistry B*, vol. 118, pp. 107–114, Jan. 2014.

[69] N. F. A. van der Vegt and D. Nayar, "The Hydrophobic Effect and the Role of Cosolvents," *The Journal of Physical Chemistry B*, vol. 121, pp. 9986–9998, Nov. 2017.

[70] A. Villa, C. Peter, and N. F. A. van der Vegt, "Transferability of Nonbonded Interaction Potentials for Coarse-Grained Simulations: Benzene in Water," *Journal of Chemical Theory and Computation*, vol. 6, pp. 2434–2444, Aug. 2010.

[71] T. Sanyal and M. S. Shell, "Transferable Coarse-Grained Models of LiquidâĂŞLiquid Equilibrium Using Local Density Potentials Optimized with the Relative Entropy," *The Journal of Physical Chemistry B*, Mar. 2018.

[72] A. Villa, C. Peter, and N. F. A. van der Vegt, "Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, p. 2077, 2009.

[73] A. Villa, N. F. A. van der Vegt, and C. Peter, "Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, p. 2068, 2009.

[74] S. P. Carmichael and M. S. Shell, "A New Multiscale Algorithm and Its Application to Coarse-Grained Peptide Models for Self-Assembly," *The Journal of Physical Chemistry B*, vol. 116, pp. 8383–8393, July 2012.

[75] J. F. Rudzinski and W. G. Noid, "Bottom-Up Coarse-Graining of Peptide Ensembles and Helix-Coil Transitions," *Journal of Chemical Theory and Computation*, vol. 11, pp. 1278–1291, Mar. 2015.

[76] Q. Sun and R. Faller, "Systematic Coarse-Graining of a Polymer Blend: Polyisoprene and Polystyrene," *Journal of Chemical Theory and Computation*, vol. 2, pp. 607–615, May 2006.

[77] B. Bayramoglu and R. Faller, "Coarse-Grained Modeling of Polystyrene in Various Environments by Iterative Boltzmann Inversion," *Macromolecules*, vol. 45, pp. 9205–9219, Nov. 2012.

[78] B. L. Peters, K. M. Salerno, A. Agrawal, D. Perahia, and G. S. Grest, "Coarse-Grained Modeling of Polyethylene Melts: Effect on Dynamics," *Journal of Chemical Theory and Computation*, vol. 13, pp. 2890–2896, June 2017.

[79] B. Hess, C. Holm, and N. van der Vegt, "Osmotic coefficients of atomistic NaCl (aq) force fields," *The Journal of Chemical Physics*, vol. 124, p. 164509, Apr. 2006.

[80] Y. Wang, W. G. Noid, P. Liu, and G. A. Voth, "Effective force coarse-graining," *Physical Chemistry Chemical Physics*, vol. 11, no. 12, p. 2002, 2009.

[81] E. Brini, V. Marcon, and N. F. A. van der Vegt, "Conditional reversible work method for molecular coarse graining applications," *Physical Chemistry Chemical Physics*, vol. 13, no. 22, p. 10468, 2011.

[82] E. Brini and N. F. A. van der Vegt, "Chemically transferable coarse-grained potentials from conditional reversible work calculations," *The Journal of Chemical Physics*, vol. 137, p. 154113, Oct. 2012.

[83] S. Izvekov and G. A. Voth, "A Multiscale Coarse-Graining Method for Biomolecular Systems," *The Journal of Physical Chemistry B*, vol. 109, pp. 2469–2473, Feb. 2005.

[84] W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, "The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models," *The Journal of Chemical Physics*, vol. 128, p. 244115, June 2008.

[85] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," *The Journal of Chemical Physics*, vol. 128, p. 244114, June 2008.

[86] F. Ercolessi and J. B. Adams, "Interatomic Potentials from First-Principles Calculations: The Force-Matching Method," *Europhysics Letters (EPL)*, vol. 26, pp. 583–588, June 1994.

[87] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, "Coarse-Grained Peptide Modeling Using a Systematic Multiscale Approach," *Biophysical Journal*, vol. 92, pp. 4289–4303, June 2007.

[88] S. Izvekov and G. A. Voth, "Multiscale Coarse-Graining of Mixed Phospholipid/Cholesterol Bilayers," *Journal of Chemical Theory and Computation*, vol. 2, pp. 637–648, May 2006.

[89] J. W. Mullinax and W. G. Noid, "Generalized-Yvon-Born-Green Theory for molecular systems," *Physical Review Letters*, vol. 103, Nov. 2009.

[90] J. W. Mullinax and W. G. Noid, "A Generalized-Yvon-Born-Green Theory for Determining Coarse-Grained Interaction Potentials," *The Journal of Physical Chemistry C*, vol. 114, pp. 5661–5674, Apr. 2010.

[91] J. F. Rudzinski and W. G. Noid, "A generalized-Yvon-Born-Green method for coarse-grained modeling: Advances, Challenges, and Insight," *The European Physical Journal Special Topics*, vol. 224, pp. 2193–2216, Sept. 2015.

[92] C. R. Ellis, J. F. Rudzinski, and W. G. Noid, "Generalized-Yvon-Born-Green Model of Toluene," *Macromolecular Theory and Simulations*, vol. 20, pp. 478–495, Aug. 2011.

[93] J. F. Rudzinski, K. Lu, S. T. Milner, J. K. Maranas, and W. G. Noid, "Extended Ensemble Approach to Transferable Potentials for Low-Resolution Coarse-Grained Models of Ionomers," *Journal of Chemical Theory and Computation*, vol. 13, pp. 2185–2201, May 2017.

[94] J. F. Rudzinski and W. G. Noid, "Investigation of Coarse-Grained Mappings via an Iterative Generalized-Yvon-Born-Green Theory Method," *The Journal of Physical Chemistry B*, vol. 118, pp. 8295–8312, July 2014.

[95] W. Schommers, "Pair potentials in disordered many-particle systems: A study for liquid gallium," *Physical Review A*, vol. 28, pp. 3599–3605, Dec. 1983.

[96] D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations: Mesoscale Potentials from Atomistic Simulations," *Journal of Computational Chemistry*, vol. 24, pp. 1624–1636, Oct. 2003.

[97] A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," *Physical Review E*, vol. 52, pp. 3730–3737, Oct. 1995.

[98] M. S. Shell, "The relative entropy is fundamental to multiscale and inverse thermodynamic problems," *The Journal of Chemical Physics*, vol. 129, p. 144108, Oct. 2008.

[99] R. A. Kullback, Solomon and Leibler, "On information and sufficiency," *Ann. Math. Stat.*, vol. 22, no. 1, pp. 79–86.

[100] M. S. Shell, "COARSE-GRAINING WITH THE RELATIVE ENTROPY," in *Advances in Chemical Physics* (S. A. Rice and A. R. Dinner, eds.), pp. 395–441, Hoboken, NJ, USA: John Wiley & Sons, Inc., Sept. 2016.

[101] J. F. Rudzinski and W. G. Noid, "Coarse-graining entropy, forces, and structures," *The Journal of Chemical Physics*, vol. 135, p. 214101, Dec. 2011.

[102] S. Riniker, J. R. Allison, and W. F. van Gunsteren, "On developing coarse-grained models for biomolecular simulation: a review," *Physical Chemistry Chemical Physics*, vol. 14, no. 36, p. 12423, 2012.

[103] M. Guenza, "Thermodynamic consistency and other challenges in coarse-graining models," *The European Physical Journal Special Topics*, vol. 224, pp. 2177–2191, Sept. 2015.

[104] H. Wang, C. Junghans, and K. Kremer, "Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining?," *The European Physical Journal E*, vol. 28, pp. 221–229, Feb. 2009.

[105] E. C. Allen and G. C. Rutledge, "A novel algorithm for creating coarse-grained, density dependent implicit solvent models," *The Journal of Chemical Physics*, vol. 128, p. 154115, Apr. 2008.

[106] J. Ghosh and R. Faller, "State point dependence of systematically coarseâĂŞgrained potentials," *Molecular Simulation*, vol. 33, pp. 759–767, Aug. 2007.

[107] P. Ganguly and N. F. A. van der Vegt, "Representability and Transferability of Kirkwood-Buff Iterative Boltzmann Inversion Models for Multicomponent Aqueous Systems," *Journal of Chemical Theory and Computation*, vol. 9, pp. 5247–5256, Dec. 2013.

[108] J. W. Mullinax and W. G. Noid, "Extended ensemble approach for deriving transferable coarse-grained potentials," *The Journal of Chemical Physics*, vol. 131, no. 10, p. 104110, 2009.

[109] T. C. Moore, C. R. Iacovella, and C. McCabe, "Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion," *The Journal of Chemical Physics*, vol. 140, p. 224104, June 2014.

[110] P. Ganguly, D. Mukherji, C. Junghans, and N. F. A. van der Vegt, "Kirkwood-Buff Coarse-Grained Force Fields for Aqueous Solutions," *Journal of Chemical Theory and Computation*, vol. 8, pp. 1802–1807, May 2012.

[111] M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*. Oxford science publications, Oxford: Clarendon Press, 2009. OCLC: 845656692.

[112] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*. No. 1 in Computational science series, San Diego: Academic Press, 2nd ed ed., 2002.

[113] M. S. Shell, *Thermodynamics and statistical mechanics: an integrated approach*. Cambridge series in chemical engineering, Cambridge: Cambridge University Press, 2015.

[114] M. E. Tuckerman, *Statistical mechanics: theory and molecular simulation*. Oxford ; New York: Oxford University Press, 2010. OCLC: ocn551495372.

[115] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. Bristol [England] ; Philadelphia: A. Hilger, special student ed ed., 1988.

[116] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of Chemical Physics*, vol. 76, pp. 637–649, Jan. 1982.

[117] L. Woodcock, "Isothermal molecular dynamics calculations for liquid salts," *Chemical Physics Letters*, vol. 10, pp. 257–261, Aug. 1971.

[118] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, pp. 3684–3690, Oct. 1984.

[119] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, pp. 255–268, June 1984.

[120] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical Review A*, vol. 31, pp. 1695–1697, Mar. 1985.

[121] H. J. C. Berendsen, *Simulating the physical world: hierarchical modeling from quantum mechanics to fluid dynamics*. 2007. OCLC: 992905220.

[122] N. Goga, A. J. Rzepiela, A. H. de Vries, S. J. Marrink, and H. J. C. Berendsen, "Efficient Algorithms for Langevin and DPD Dynamics," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3637–3649, Oct. 2012.

[123] S. Nosé and M. Klein, "Constant pressure molecular dynamics for molecular systems," *Molecular Physics*, vol. 50, pp. 1055–1076, Dec. 1983.

[124] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé-Hoover chains: The canonical ensemble via continuous dynamics," *The Journal of Chemical Physics*, vol. 97, pp. 2635–2643, Aug. 1992.

[125] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein, "Explicit reversible integrators for extended systems dynamics," *Molecular Physics*, vol. 87, pp. 1117–1157, Apr. 1996.

[126] N. Zacharopoulos, N. Vergadou, and D. N. Theodorou, "Coarse graining using pretabulated potentials: Liquid benzene," *The Journal of Chemical Physics*, vol. 122, p. 244111, June 2005.

[127] B. M. Mognetti, L. Yelash, P. Virnau, W. Paul, K. Binder, M. Müller, and L. G. MacDowell, "Efficient prediction of thermodynamic properties of quadrupolar fluids from simulation of a coarse-grained model: The case of carbon dioxide," *The Journal of Chemical Physics*, vol. 128, p. 104501, Mar. 2008.

[128] R. Henderson, "A uniqueness theorem for fluid pair correlation functions," *Physics Letters A*, vol. 49, pp. 197–198, Sept. 1974.

[129] S. Jain, S. Garde, and S. K. Kumar, "Do Inverse Monte Carlo Algorithms Yield Thermodynamically Consistent Interaction Potentials?," *Industrial & Engineering Chemistry Research*, vol. 45, pp. 5614–5618, Aug. 2006.

[130] C.-C. Fu, P. M. Kulkarni, M. Scott Shell, and L. Gary Leal, "A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties," *The Journal of Chemical Physics*, vol. 137, p. 164106, Oct. 2012.

[131] D. Rosenberger, M. Hanke, and N. F. van der Vegt, "Comparison of iterative inverse coarse-graining methods," *The European Physical Journal Special Topics*, vol. 225, pp. 1323–1345, Oct. 2016.

[132] D. Ivanizki, *Numerical analysis of the relation between interactions and structure in a molecular fluid*. Mainz: Univ., 2015.

[133] T. T. Foley, M. S. Shell, and W. G. Noid, "The impact of resolution upon entropy and information in coarse-grained models," *The Journal of Chemical Physics*, vol. 143, p. 243104, Dec. 2015.

[134] M. E. Johnson, T. Head-Gordon, and A. A. Louis, "Representability problems for coarse-grained water potentials," *The Journal of Chemical Physics*, vol. 126, p. 144509, Apr. 2007.

[135] T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen, "Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints," *The Journal of Chemical Physics*, vol. 126, p. 075101, Feb. 2007.

[136] A. Moradzadeh, M. H. Motevaselian, S. Y. Mashayak, and N. R. Aluru, "Coarse-Grained Force Field for Imidazolium-Based Ionic Liquids," *Journal of Chemical Theory and Computation*, vol. 14, pp. 3252–3261, June 2018.

[137] F. Delbary, M. Hanke, and D. Ivanizki, "A generalized Newton iteration for computing the solution of the inverse Henderson problem," *arXiv:1806.11135 [math]*, June 2018. arXiv: 1806.11135.

[138] N. G. Almarza and E. Lomba, "Determination of the interaction potential from the pair distribution function: An inverse Monte Carlo technique," *Physical Review E*, vol. 68, July 2003.

[139] N. Ashcroft and D. Stroud, "Theory of the Thermodynamics of Simple Liquid Metals," in *Solid State Physics*, vol. 33, pp. 1–81, Elsevier, 1978.

[140] E. Canessa, M. Silbert, and M. Grimson, "Volume dependent forces in charge stabilized colloidal crystals," *Molecular Physics*, vol. 64, pp. 1195–1201, Aug. 1988.

[141] G. C. Barker, M. J. Grimson, and M. Silbert, "Effective interactions in colloidal dispersions stabilized by free polymers," *Molecular Physics*, vol. 84, pp. 211–215, Jan. 1995.

[142] A. A. Louis, "Beware of density dependent pair potentials," *Journal of Physics: Condensed Matter*, vol. 14, pp. 9187–9206, Oct. 2002.

[143] A. Das and H. C. Andersen, "The multiscale coarse-graining method. V. Isothermal-isobaric ensemble," *The Journal of Chemical Physics*, vol. 132, p. 164106, Apr. 2010.

[144] T. Sanyal and M. S. Shell, "Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation," *The Journal of Chemical Physics*, vol. 145, p. 034109, July 2016.

[145] N. J. H. Dunn and W. G. Noid, "Bottom-up coarse-grained models with predictive accuracy and transferability for both structural and thermodynamic properties of heptane-toluene mixtures," *The Journal of Chemical Physics*, vol. 144, p. 204124, May 2016.

[146] D. Rosenberger and N. F. A. van der Vegt, "Addressing the temperature transferability of structure based coarse graining models," *Physical Chemistry Chemical Physics*, vol. 20, no. 9, pp. 6617–6628, 2018.

[147] M. R. DeLyser and W. G. Noid, "Extending pressure-matching to inhomogeneous systems via local-density potentials," *The Journal of Chemical Physics*, vol. 147, p. 134111, Oct. 2017.

[148] J. F. Rudzinski, K. Lu, S. T. Milner, J. K. Maranas, and W. G. Noid, "Extended Ensemble Approach to Transferable Potentials for Low-Resolution Coarse-Grained Models of Ionomers," *Journal of Chemical Theory and Computation*, vol. 13, pp. 2185–2201, May 2017.

[149] J. F. Dama, J. Jin, and G. A. Voth, "The Theory of Ultra-Coarse-Graining. 3. Coarse-Grained Sites with Rapid Local Equilibrium of Internal States," *Journal of Chemical Theory and Computation*, vol. 13, pp. 1010–1022, Mar. 2017.

[150] K. Farah, A. C. Fogarty, M. C. Böhm, and F. Müller-Plathe, "Temperature dependence of coarse-grained potentials for liquid hexane," *Phys. Chem. Chem. Phys.*, vol. 13, no. 7, pp. 2894–2902, 2011.

[151] I. M. Ilie, W. K. den Otter, and W. J. Briels, "A coarse grained protein model with internal degrees of freedom. Application to *Îś* -synuclein aggregation," *The Journal of Chemical Physics*, vol. 144, p. 085103, Feb. 2016.

[152] J. W. Wagner, T. Dannenhoffer-Lafage, J. Jin, and G. A. Voth, "Extending the range and physical accuracy of coarse-grained models: Order parameter dependent interactions," *The Journal of Chemical Physics*, vol. 147, p. 044113, July 2017.

[153] J. Jin and G. A. Voth, "Ultra-Coarse-Grained Models Allow for an Accurate and Transferable Treatment of Interfacial Systems," *Journal of Chemical Theory and Computation*, vol. 14, pp. 2180–2197, Apr. 2018.

[154] N. J. H. Dunn, T. T. Foley, and W. G. Noid, "Van der Waals Perspective on Coarse-Graining: Progress toward Solving Representability and Transferability Problems," *Accounts of Chemical Research*, vol. 49, pp. 2832–2840, Dec. 2016.

[155] M. A. Webb, J.-Y. Delannoy, and J. J. de Pablo, "Graph-Based Approach to Systematic Molecular Coarse-Graining," *Journal of Chemical Theory and Computation*, Dec. 2018.

[156] W. Wang and R. Gómez-Bombarelli, "Variational Coarse-Graining for Molecular Dynamics," *arXiv:1812.02706 [physics, stat]*, Dec. 2018. arXiv: 1812.02706.

[157] T. Dannenhoffer-Lafage, A. D. White, and G. A. Voth, "A Direct Method for Incorporating Experimental Data into Multiscale Coarse-Grained Models," *Journal of Chemical Theory and Computation*, vol. 12, pp. 2144–2153, May 2016.

[158] M. H. Motevaselian, S. Y. Mashayak, and N. R. Aluru, "Extended coarse-grained dipole model for polar liquids: Application to bulk and confined water," *Physical Review E*, vol. 98, Nov. 2018.

[159] J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, "The Theory of Ultra-Coarse-Graining. 1. General Principles," *Journal of Chemical Theory and Computation*, vol. 9, pp. 2466–2480, May 2013.

[160] T. Bereau and J. F. Rudzinski, "Accurate Structure-Based Coarse Graining Leads to Consistent Barrier-Crossing Dynamics," *Physical Review Letters*, vol. 121, Dec. 2018.

[161] K. K. Bejagam, S. Singh, Y. An, and S. A. Deshmukh, "Machine-Learned Coarse-Grained Models," *The Journal of Physical Chemistry Letters*, vol. 9, pp. 4667–4672, Aug. 2018.

[162] K. K. Bejagam, Y. An, S. Singh, and S. A. Deshmukh, "Machine-Learning Enabled New Insights into the Coil-to-Globule Transition of Thermosensitive Polymers Using a Coarse-Grained Model," *The Journal of Physical Chemistry Letters*, vol. 9, pp. 6480–6488, Nov. 2018.

[163] S. T. John and G. Csányi, "Many-Body Coarse-Grained Interactions Using Gaussian Approximation Potentials," *The Journal of Physical Chemistry B*, vol. 121, pp. 10934–10949, Dec. 2017.

[164] R. L. C. Akkermans and W. J. Briels, "Coarse-grained dynamics of one chain in a polymer melt," *The Journal of Chemical Physics*, vol. 113, pp. 6409–6422, Oct. 2000.

[165] S. Izvekov and G. A. Voth, "Modeling real dynamics in the coarse-grained representation of condensed phase systems," *The Journal of Chemical Physics*, vol. 125, p. 151101, Oct. 2006.

[166] C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni, "MorinZwanzig formalism as a practical computational tool," *Faraday Discuss.*, vol. 144, pp. 301–322, 2010.

[167] D. Fritz, C. R. Herbers, K. Kremer, and N. F. A. van der Vegt, "Hierarchical modeling of polymer permeation," *Soft Matter*, vol. 5, no. 22, p. 4556, 2009.

[168] V. A. Harmandaris and K. Kremer, "Dynamics of Polystyrene Melts through Hierarchical Multi-scale Simulations," *Macromolecules*, vol. 42, pp. 791–802, Feb. 2009.

[169] R. Zwanzig, "Memory Effects in Irreversible Thermodynamics," *Physical Review*, vol. 124, pp. 983–992, Nov. 1961.

[170] H. Mori, "Transport, Collective Motion, and Brownian Motion," *Progress of Theoretical Physics*, vol. 33, pp. 423–455, Mar. 1965.

[171] R. D. Groot and P. B. Warren, "Dissipative particle dynamics: Bridging the gap between atomistic and mesoscopic simulation," *The Journal of Chemical Physics*, vol. 107, pp. 4423–4435, Sept. 1997.

[172] P. Español and P. B. Warren, "Perspective: Dissipative particle dynamics," *The Journal of Chemical Physics*, vol. 146, p. 150901, Apr. 2017.

[173] A. Eriksson, M. N. Jacobi, J. Nyström, and K. Tunstrøm, "A method for estimating the interactions in dissipative particle dynamics from particle trajectories," *Journal of Physics: Condensed Matter*, vol. 21, p. 095401, Mar. 2009.

[174] Z. Li, X. Bian, B. Caswell, and G. E. Karniadakis, "Construction of dissipative particle dynamics models for complex fluids via the Mori-Zwanzig formulation," *Soft Matter*, vol. 10, no. 43, pp. 8659–8672, 2014.

[175] H. Lei, B. Caswell, and G. E. Karniadakis, "Direct construction of mesoscopic models from microscopic simulations," *Physical Review E*, vol. 81, Feb. 2010.

[176] Z. Li, X. Bian, X. Li, and G. E. Karniadakis, "Incorporation of memory effects in coarse-grained modeling via the Mori-Zwanzig formalism," *The Journal of Chemical Physics*, vol. 143, p. 243128, Dec. 2015.

[177] Z. Li, H. S. Lee, E. Darve, and G. E. Karniadakis, "Computing the non-Markovian coarse-grained interactions derived from the MoriâĂŞZwanzig formalism in molecular systems: Application to polymer melts," *The Journal of Chemical Physics*, vol. 146, p. 014104, Jan. 2017.

[178] A. Eriksson, M. N. Jacobi, J. Nyström, and K. Tunstrøm, "Effective thermostat induced by coarse graining of simple point charge water," *The Journal of Chemical Physics*, vol. 129, p. 024106, July 2008.

[179] A. Eriksson, M. N. Jacobi, J. Nyström, and K. Tunstrøm, "Bottom-up derivation of an effective thermostat for united atoms simulations of water," *The Journal of Chemical Physics*, vol. 130, p. 164509, Apr. 2009.

[180] S. Izvekov and B. M. Rice, "Multi-scale coarse-graining of non-conservative interactions in molecular liquids," *The Journal of Chemical Physics*, vol. 140, p. 104104, Mar. 2014.

[181] S. Trément, B. Schnell, L. Petitjean, M. Couty, and B. Rousseau, "Conservative and dissipative force field for simulation of coarse-grained alkane molecules: A bottom-up approach," *The Journal of Chemical Physics*, vol. 140, p. 134113, Apr. 2014.

[182] C. A. Lemarchand, M. Couty, and B. Rousseau, "Coarse-grained simulations of *cis* - and *trans* -polybutadiene: A bottom-up approach," *The Journal of Chemical Physics*, vol. 146, p. 074904, Feb. 2017.

[183] G. Deichmann and N. F. A. van der Vegt, "Bottom-up approach to represent dynamic properties in coarse-grained molecular simulations," *The Journal of Chemical Physics*, vol. 149, p. 244114, Dec. 2018.

[184] F. Noé, "Probability distributions of molecular observables computed from Markov models," *The Journal of Chemical Physics*, vol. 128, p. 244103, June 2008.

[185] J. F. Rudzinski, K. Kremer, and T. Bereau, "Communication: Consistent interpretation of molecular simulation kinetics using Markov state models biased with external information," *The Journal of Chemical Physics*, vol. 144, p. 051102, Feb. 2016.

[186] J. Rudzinski and T. Bereau, "Concurrent parametrization against static and kinetic information leads to more robust coarse-grained force fields," *The European Physical Journal Special Topics*, vol. 225, pp. 1373–1389, Oct. 2016.